의학박사 학위논문

Genomic and transcriptomic landscape of two-step malignant transformation in x-irradiated C3H10T1/2 cells

X-선 조사 C3H10T1/2 세포주에서 단계별 형질전환의 유전체 및 전사체 연구

2019년 2월

서울대학교 대학원

의과학과 의과학전공

조 벼 리

X-선 조사 C3H10T1/2 세포주에서

단계별 형질전환의 유전체 및 전사체 연구

지도교수 김 종 일

이 논문을 의학박사 학위논문으로 제출함

2019년 2월

서울대학교 의과대학원

의과학과 의과학전공

조 벼 리

조벼리의 의학박사 학위논문을 인준함

2018년 12월

위 원 장 윤 홍 덕 Hong Duk Youn (인)

부위원장 김 종 일 (인)

위 원 서 지 선 (인)

위 원 조 성 엽 (인)

위 원 안 정 현 (인)

Genomic and transcriptomic landscape of two-step malignant transformation in x-irradiated C3H10T1/2 cells

by

Byuri Cho

A thesis submitted to the Department of Biomedical Science in partial fulfillment of the requirement of the Degree of Doctor of Philosophy in Medical Science at Seoul National University College of Medicine

December 2018

Approved by Thesis Committee:

Professor Hongdeok Youn Chairman

Professor JONG-IL KIM Vice chairman

Professor Jeong Suk Seo.

Professor Sung-Yup Cho

Professor Sung Hyuck AHN

ABSTRACT

Genomic and transcriptomic landscape of two-step malignant transformation in x-irradiated C3H10T1/2 cells

Byuri Cho
Major in Biomedical Sciences
Seoul National University College of Medicine
Seoul National University Graduate School

Tumourigenesis is initiated by various factors and its progression is also affected by numerous elements. In 1980, Little and Kennedy performed malignant cell transformation experiment using x-irradiation to study the initiation and progression of tumourigenesis. Reproducing tumourigenesis *in vitro* is challenging, yet, they accomplished the task and almost 40 years later, I am here to elucidate unsolved questions beneath that experiment. Firstly, genomic alterations were investigated

using whole-genome sequencing. Although malignant transformation process is a recreation of tumourigenesis *in vitro*, the genomic alterations were not as striking as I expected which was very intriguing. The generation of focus definitely implied the presence of DNA damage, yet the changes occurred at this stage were not quite evolved. I then studied transcriptome changes that arose due to x-irradiation. Using principal component analysis and admixture analysis, I illustrated that the transcriptome profiles were distinctively differed according to x-irradiation or focus generation status. To investigate what drives focus generation, I analysed differentially expressed genes between un-irradiated cells and irradiated but non-focus cells. Non-focus cells seemed to be at the initial stage of malignant transformation as its characteristics were closer to focus rather than un-irradiated cells. Particularly, non-focus cells exhibited highly elevated *Cdkn1a*, a DNA damage response gene, possibly to respond to DNA damage by x-irradiation. Also, down-regulation of TGF-β genes were observed in non-focus cells which may be one of the factors that induce focus generation. Moreover, DNA repair related genes including *Atm*, *Atr*, *Brca1*, *Brca2*, and *Chek1* were highly elevated in focus cells. This study described that alterations due to x-irradiation in transcriptome were quite dramatic whereas changes in genome was rather gentle. Furthermore, as stem cells are frequently involved in malignant transformation, whether Focus possesses stem cell-like characteristics was examined. As a result, Focus displayed “stemness” and up-regulated *Myc* contributed to oncogenic reprogramming hence generated Focus cells with tumourigenic characteristics. In conclusion, I speculated that there are three factors involved in the second step of two-step malignant transformation. The

primary factor would be down-regulated *Tgfβ* gene expression. The change occurred in Non-focus, which is considered as at the very early stage of tumourigenesis. Decreased *Tgfβ* gene expression led the cells to transform further, hence induced DNA repair process especially error-prone DNA repair which would act as the secondary factor. Lastly, with up-regulated *Myc* oncogene, the cells were reprogrammed into cancer stem cell, hence generated foci with malignancy. Also, Focus cells exhibited stem cell-like characteristics. Taken together, in this study, I uncovered three factors that contributed to the second step of two-step malignant transformation and I believe that these findings will extend the understanding of tumourigenesis.

Keywords: Malignant transformation; Tumourigenesis; RNA sequencing; Whole-genome sequencing; X-irradiation; TGF-β genes; Error-prone DNA repair pathway; Cancer stem cell; Oncogenic reprogramming

Student number: 2015-31237

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CGP: Chemical and genetic perturbations

Chr: Chromosome

COSMIC: Catalogue of somatic mutations in cancer

DEG: Differentially expressed genes

DSB: Double-strand break

E. coli: Escherichia coli

ESC: Embryonic stem cell

GATK: Genome analysis toolkit

GTEx: Genotype-tissue expression project

GSEA: Gene set enrichment analysis

HM: Hallmark

HR: Homologous recombination

KEGG: Kyoto encyclopedia of genes and genomes

MMP: Matrix metalloproteinase

NGS: Next-generation sequencing

NHEJ: Non-homologous end-joining

PCA: Principal component analysis

PCR: Polymerase chain reaction

RNA-seq: RNA sequencing

SCNV: Somatic copy number variation

SNP: Single nucleotide polymorphism

SV: Structural variation

TMB: Tumour mutation burden

UV: Ultraviolet

VST: Variance stabilising transformation

WGCNA: Weighted gene co-expression analysis

WGS: Whole-genome sequencing

INTRODUCTION

History of studying tumourigenesis

Cancer occurs in various forms with numerous causes. Sometimes one significant driver gene mutation initiates tumourigenesis, but most cases, accumulated multiple mutations generate tumours [1]. For many decades, there had been many attempts to solve the underlying mechanism of tumourigenesis. However, cancer still is the disease which threats human life the most.

In the modern society, the life expectancy is dramatically increasing and prolonged lifetime drew more attention to cure of lethal diseases such as cancer in order to find a way to enhance the quality of health. There are various causes of lethal illnesses. Aging is the most common factor for initiating fetal diseases as one always go through physiological changes as part of natural life. It is a gradual and permanent process although not a pathological event. Aging does cause many disease including cancer, however, the most frequent cause of cancer is spontaneous mutations in genes that play a crucial role in tumourigenesis [2].

Cancer arises by somatic mutations in cancer driver genes such as oncogenes and tumour suppressor genes. For many years, the field of cancer research tried to find the causation of cancer. The very first insights on human cancer initiation was arisen in the 18th century [3]. From that time point, many intriguing thoughts and theories on tumourigenesis initiation were followed and in the early 20th century, Yamagiwa and Ichikawa reported their study on multistage and multifactor nature of

the tumourigenesis process using experimental technique [4]. Also, the first cancer virus was discovered in chickens by Ellerman and Bang [5] and Rous [6]. Then, a few decades later, Rous expanded his earlier studies on cancer virus and identified *src* oncogene, which is the first oncogene ever to be discovered in the history of cancer biology [6, 7].

Many hypotheses regarding finding the causes and consequences of tumourigenesis had been suggested. In the beginning, majority of cancer studies were focused on cytology and genetics using cellular and mouse model experiments. They were followed by researches that highlighted the importance of epigenetically controlled gene expression and genomic alterations in cancer to see the bigger picture and broaden the understanding of tumourigenesis. John Little and Ann Kennedy also proposed a hypothesis that there were two steps that contributed to tumour initiation and development. They believed that in the first step, an element from external environment, x-irradiation, induced changed characteristics of cells, then the second element altered the features of cells completely including their morphology. They called this hypothesis as "two-step hypothesis" and they provided some evidence by conducting cellular experiments [8].

Two-step hypothesis of malignant cell transformation

It is known that the accumulation of multiple genetic alterations is required for malignant transformation of a cell during tumour initiation or progression. In 1976, Little conducted an experiment to study x-irradiation induced oncogenic transformation in C3H10T1/2 cell line [9]. The foci were generated upon x-

irradiation exposure and their oncogenic effect was confirmed using mouse model. Firstly, Little categorised formed foci according to their morphological features. Type 1 cells only showed unusual growth pattern but no obvious defected morphology. Type 2 and Type 3 were counted as transformed cells as Type 2 cells exhibited stellate features and most of Type 3 cells were piled up and formed multi-layers of cells. The oncogenic effect of Type 2 and Type 3 focus was confirmed when tumour incidence was observed upon their injection. Tumours were formed when type 2 or type 3 focus cells were injected, not type 1. Then, a few years later, Little further investigated x-irradiation induced malignant transformation in C3H10T1/2 cells [8, 10]. C3H10T1/2 cells were exposed to various intensities of x-irradiation, then cultured at optimum condition until confluence was reached. C3H10T1/2 cells were used as they are highly sensitive to post-confluence inhibition of cells division, therefore suitable cell line for transformation study. Average number of one focus per dish was generated after going through 12-13 rounds of cell division. When non-focus cells from x-irradiated dish were re-seeded, they also formed focus after similar numbers of cell division at confluence. When normal cells reach at confluency, they become density inhibited. Transformed cells, however, display no sign of density inhibition which is one of the characteristics of transformed cells. It was very intriguing result that the focus was formed after going through certain numbers of cell division, not right after x-irradiation exposure. Based on this result, Little proposed that there were at least two steps involved in malignant transformation process. The hypothesis suggested by Little and Kennedy described that the initial damage was occurred due to x-irradiation, and then unknown element,

acted as a second hit, induced malignant transformation process. Therefore, Little's experiments may be considered as an illustration of indirect tumour induction after x-irradiation. There were several subsequent studies regarding two-step malignant transformation including Little's another investigation and Park *et al.*'s research [11]. In both studies, they tried to elucidate the factors that contributed to the second step of two-step malignant transformation. However, no clear answer was discovered at that time.

Error-free and error-prone DNA repair system

There are several reported studies suggesting defects in DNA repair system may increase the incidence of tumourigenesis [12]. In *Escherichia coli* (*E. coli*), when error-free repair system is saturated, as a last resort, error-prone repair system is activated and it is called the SOS system of bacteria [13-15]. The SOS system is usually responsible for ultraviolet (UV) induced mutations [14, 16], and the yield of DNA repair is quite low. This error-prone repair system may function in mutagenesis hence carcinogenesis. Here, I can speculate that Little's experiment reflected that of DNA damaged *E. coli*. The x-irradiation firstly damaged DNA then altered some cellular processes which induced error-prone repair system, hence slowly generated transformants. In 1980s, the whole process was obscure and vague but I was able to anticipate that error-prone repair system in mammalian cell that is similar to the SOS system existed and contributed to malignant transformation [17]. I believe that x-irradiation indirectly turned on the error-prone repair system and promoted the generation of transformants. Unfortunately, without having full knowledge on gene

expression profile or genomic alterations, some parts remained unanswered.

Several decades later, in this study, I performed further analysis to give answers to unsolved questions using next-generation sequencing (NGS) technology. In the past, with experimental based studies, it was quite challenging to elucidate the causes and consequences of tumourigenesis. However, in recent years, various techniques that can identify genomic and transcriptomic characteristics of cancer. With NGS method, vast amount of information could be found and utilised to study tumourigenesis. The NGS techniques have definitely expanded the field of cancer biology and even help to discover therapeutic targets of cancer.

Application of next-generation sequencing

Next-generation sequencing, also known as massively parallel or deep sequencing, is now considered as a gold standard method in the field of cancer genomics which gives faster and more accurate results when compared to traditional Sanger sequencing [18-20]. This method revolutionised the whole field of genomic research. The traditional Sanger sequencing is considered as the first-generation sequencing technique, and NGS, the second-generation sequencing, delivers vast amount of high throughput data with reduced expense. Thus, this extraordinary technology has opened a new chapter of cancer biology and cancer genomics.

The applications of NGS are diverse and new systems are being developed continually. These applications can be categorised based on the purpose of the study.

(1) To build a brand-new reference genome from known or unknown

organism, *de novo* sequencing technique may be applied. Usually, *de novo* sequencing denotes to sequencing a unique genome where there is no reference sequence accessible for alignment. However, to improve the quality of result and to discover population specific genetic variation, the reference genome can be rebuilt. In 2016, Seo *et al*. generated AK1 genome reference which is based on Korean human genome [21]. Although human genome reference already exists, the need for population specific reference always remained. The establishment of AK1 human genome reference fulfilled these needs by providing more accurate result of genetic variation for Korean patients.

(2) To study human genetic variation, whole-genome sequencing (WGS) or whole-exome sequencing (WES) techniques can be applied using existing reference genome [22, 23]. Obtained sequencing results are compared with reference genome, then researchers can find the genetic variations including single-nucleotide polymorphisms (SNPs), structural variations, and somatic copy number alterations.

(3) To elucidate the gene expression pattern, RNA sequencing (RNA-seq) can be practiced. RNA-seq is a ground-breaking tool for transcriptome profiling [24]. It allows researchers to quantify gene expression level and characterise alternative splicing. Also, identifying novel transcript can be done using this particular method. Above all, the most appealing feature of RNA-seq is that the expression level of each gene can be compared and studied between groups of samples under various conditions [25-27].

(4) For epigenetic studies, DNA methylation may be analysed, and

sequencing methods such as bisulfite sequencing can be used. To determine methylation of CpG island, this technique is needed. Bisulfite treatment converts unmethylated cytosine to uracil so that only the methylated cytosine residues are identified. As altered methylation pattern is often observed in various types of human cancers, this tool offers valuable results that can characterise the DNA methylation profile of cancer cells [28, 29].

As mentioned above, NGS techniques hold many advantages including providing large amount of high throughput data as well as cost reduction. As the technique offers great extent of data, diverse genomic variations and gene expression patterns can be studied. The technique can be applied in the field of clinical research such as diagnosis of rare disease using molecular markers, non-invasive parental diagnosis, and of course, cancer biology [30, 31].

Study scheme

Lately, with greatly advanced NGS technology, many unsolved questions now have answers to them. Here, I reproduced Little's experiment, then performed in-depth analysis using WGS and RNA-seq method to elucidate the underlying causes of malignant transformation. The samples were grouped into three categories, un-irradiated cells, irradiated but non-focus cells, and irradiated focus cells. Then those samples were subjected to genomic and transcriptomic analyses.

In the present study, I described the effect of x-irradiation in initiation and progression of malignant transformation, in other words, tumourigenesis. It is reported that radiation can induce genomic instability in mammalian cells [32]. The

consequences of radiation-induced genomic instability include chromosomal rearrangements and gene expression changes [33]. Although it is reported that radiation does not directly initiate genomic instability, but still induces defective cellular response which leads to DNA damage or gene expression alteration, hence disturbs cellular homeostasis [33-35]. Here, I examined the genomic and transcriptomic landscapes of two-step malignant transformation in x-irradiated C3H10T1/2 cells to discover the factors that contributed to the second step of malignant transformation. I believe that this study will improve our understanding of tumourigenesis.

MATERIALS AND METHODS

Cell culture

C3H10T1/2 cell line was purchased from ATCC (Cat. No. CCL-226; VA, USA). The cells were cultivated in Eagle's Basal medium with 2mM L-glutamine, 1mM sodium pyruvate, and 1500mg/L sodium bicarbonate (Cat. No. 30-2003; ATCC, VA, USA) containing 10% fetal bovine serum (Cat. No. 30-2020; ATCC, VA, USA) and 1% penicillin (Cat. No. SV30010; HyClone, UT, USA) and incubated at 37℃ in 5% CO_2. The cells were counted using hemocytometer and approximately 4.0 x 10^5 cells were seeded in 100mm culture dish (Falcon 100mm TC-treated cell culture dish, Cat. No. 353003; Corning, NY, USA). Upon focus formation, the cells that were subjected to RNA-seq and WGS were obtained using cylinder isolation technique (Scienceware cloning cylinders, Cat. No. C3858-50EA; Sigma, MO, USA) [9, 36].

X-irradiation

C3H10T1/2 cells were exposed to 400rads, 600rads, and 800rads of x-ray at 300cGy/min dose rate. The x-irradiation was directly exposed to culture dish. Varian 6EX linear accelerator (Clinac iX system linear accelerator; Varian, CA, USA) was used for x-irradiation.

Cell imaging

The cell images were taken before x-irradiation and after focus formation (Leica DFC490; Leica Microsystems, Wetzlar, Germany). The magnification for the cell image was x100.

Giemsa staining

To confirm focus formation giemsa staining was performed. Briefly, the cells were washed with 1X PBS (Cat. No. LB004-01; Welgen, Korea), then fixed with 4% paraformaldehyde solution (Cat. No. P2031; Biosesang, Korea) at room temperature for 30min. Next, diluted giemsa stain solution (Cat. No. 32884-250ML; Sigma, MO, USA) was placed on the culture dish at room temperature. After 2hrs, the dish was washed with 1X PBS then air dried. Blue stain represented focus. The images were taken by Cannon EF 50mm F1.2L USM camera.

RNA sequencing

RNA was extracted from each group of cells, non-irradiated, irradiated non-focus cells, and irradiated focus cells, using Easy Spin RNA extraction kit (Intron, Daejeon, Korea). Extracted RNA underwent quality and concentration measurement using NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Then, the RNA-seq was performed on a HiSeq 2000 platform (Illumina, San Diego, CA, USA). The sequenced paired-end reads were aligned to mm10 mouse reference genome using STAR 1-pass method [37, 38], then PCR duplicates were removed by using Picard MarkDuplicate (http://picard.sourceforge.net). The filtered reads underwent further processing for variant calling using the best-practice of GATK (https://software.broadinstitute.org/gatk/best-practices) [39, 40]. The summary of sequencing throughput and alignment yield is described on Table 5.

Whole-genome sequencing

Extraction of DNA was performed using QIAamp DNA mini kit (Cat. No. 51304; Qiagen, Hilden, Germany). WGS using Hiseq 2000 sequencing platform (Illumina, San Diego, CA, USA) was performed according to manufacturer's instruction. Paired-end reads were aligned to GRCh37.p13 reference using BWA [41]. The sequencing summaries of WGS are provided in Table 2. Duplicated reads were removed by Picard tools (https://broadinstitute.github.io/picard/). MuTect was used for somatic SNV calling [42]. Somatic SNVs with PASS flag were annotated by ANNOVAR [43]. To identify somatic translocations, we used CNVKit [44].

Differentially expressed gene analysis

Based on Ensemble gene set, the number of reads aligned to each gene were counted using HTSeq-count then normalized via regularized log transformation using DESeq2 [26, 45]. Here, differentially expressed genes (DEGs) were determined by DESeq2 with q-value < 0.05, $|\text{Log}_2 \text{(fold-change)}| \geq 0.6$, and baseMean ≥ 100.

Admixture analysis

To examine the potential structure of gene expression data by clustering genes CountClust (version 1.3.1) was used [46, 47]. The variance stabilizing transformed expression values from DESeq2 were used for this analysis following the developer's instructions [26, 48]. The number of genes that consisted each cluster was 1,000 and with these genes the samples were partitioned into two distinctive groups.

Weighted correlation network analysis

WGCNA was used to identify the relationship between gene expression changes in samples with focus generation status [49]. The detailed methodology is described in our previous study [50]. In brief, pairwise correlations between the expression of each gene were used to generate modules which represent the co-expression networks. To identify the associations of the constructed modules and the clinical traits, the correlation between modules and clinical features were defined using the Pearson correlation coefficient. The variance stabilizing transformed expression values from DESeq2 were used for the analysis [26, 48].

Pathway enrichment analysis

The pathway enrichment analysis, which estimates biologically significant pathways using overlaps of genes of interest with particular gene sets or pathways, was performed using the Molecular Signature Database (MSigDB) version 6.0. Genes that were involved in differentiating focus cells from non-focus cells and control cells from irradiated non-focus cells were subjected to the analysis in addition to genes that contributed in distinguishing characteristics of focus cells and non-focus cells by constructing statistically significant modules in WGCNA. Two databases from MSigDB were utilised for the analysis: the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Hallmark (HM) gene set [51, 52]. KEGG is an important source for comprehending biological systems, organisms, and even the ecosystem. The database was constructed from combined data of genome sequencing and high-throughput experiment results. HM gene sets encapsulated and

characterised precise biological status or procedures and the gene sets were created using computational methodology.

Somatic copy number variations

To observe somatic copy number alterations CNVkit was used. CNVkit calculated and visualised copy number of samples from DNA sequencing data using Python library and command-line [44, 53, 54].

Structural variations

Delly2 computed structural variants using integrated prediction method. Variations such as deletions, duplications, inversions and translocations at single-nucleotide level from sequencing data can be discovered [55].

Mutational signature analysis

Somatic mutational signatures were identified using SomaticSignature R package [56]. This R package delivered accurate and efficient form of the somatic motifs according to a record of somatically mutated genomic sites and the prediction of somatic signatures with mathematical algorithms. The estimated mutational signatures from study subjects were compared with previously defined signatures on COSMIC to decipher the characteristics of discovered signatures [57].

RESULTS

X-irradiation altered cell morphology and induced transformation of C3H10T1/2 cells

The x-irradiation induced malignant cell transformation experiment was reproduced using similar protocol from Kennedy *et al.* [8] (Figure 1). C3H10T1/2 cells were exposed to different intensities of x-irradiation, 400rads, 600rads, and 800rads, then cultured until confluency. After numerous numbers of cell divisions, foci were formed. The number of focus formed was not highly affected by the intensity of x-irradiation (Figure 2A), however, the time taken until focus formation was reduced for re-seeded (2^{nd} seeding) cells (Figure 2B). The formation of focus was observed under microscope (Figure 3A) then confirmed with Giemsa staining (Figure 3B). The morphology of transformed cells was quite differed from that of normal. The cells were no longer growing in monolayer but piled up and displayed dense growth with stellate characteristics [9]. In previous study, Little categorised foci into three groups according to the morphological characteristics. Type 1 was not considered as focus but Type 2 and 3 were scored as focus as they exhibited significantly disrupted morphology including stellate shape [9, 11]. In this study, majority of observed foci were found as type 3 focus (Figure 3C).

In 1980, Little and Kenney proposed two-step hypothesis and suggested that x-irradiation induced malignant transformation is two-step process. When this experiment was firstly performed almost 40 years ago, not sufficient technology was available to further elucidate the genomic and transcriptomic characteristics. Several

decades later, in this study, using next-generation sequencing method, I identified the genomic and transcriptomic characteristics of x-irradiated non-focus cells and transformed focus cells hence uncovered the second factor of two-step hypothesis which Little originally proposed.

Somatic copy number variation induced by x-irradiation

To observe genomic changes due to x-irradiation, WGS was performed (Table 1 and Table 2). For genomic analysis, 14 samples were subjected including two un-irradiated cells (hereafter referred as Control), four irradiated but non-focus cells (hereafter referred as Non-focus), and eight irradiated focus cells (hereafter referred as Focus). In human cancers, there usually are a few driver mutations that initiated cancer development and various types of chromosome rearrangements or copy number changes are observed [58, 59]. As this experiment mimicked malignant transformation *in vitro*, I expected a small number of driver mutations as well as chromosome aberrations to be found. However, genomic alterations induced by x-irradiation were very mild. In case of somatic copy number variation (SCNV), the alteration patterns were not easily distinguishable according to the focus status (Figure 4). Non-focus group exhibited slightly less deletion or duplication of chromosome segments when compared to Focus but the SCNV patterns were very similar to each other. In mammalian cells, SCNV functions importantly in generating essential variations in the population as well as disease phenotype [60]. Here, x-irradiation did not initiate dramatic changes in genome in terms of copy number variation. Gentle genomic alterations might suggest that focus generation may be

affected by other factors such as alterations in transcriptome [61, 62].

Tumour mutation burden of x-irradiated cells

In recent years, the prevalence of somatic mutation across human cancers was measured to elucidate the characteristics of diverse types of cancers [57]. Even though Non-focus and Focus cells are not fully developed tumour cells, as they presented tumourigenic features, tumour mutation burden (TMB) was measured. In human cancer, depends on cancer type, TMB varied. Some cancers displayed high prevalence of somatic mutations such as melanoma and lung squamous cancer and some other cancer exhibited comparatively low prevalence for example thyroid cancer and pilocytic astrocytoma. The number of somatic mutation found from Non-focus and Focus samples were extremely low when compared to human cancer samples and when Non-focus and Focus were compared, the TMB were quite similar to each other (Figure 5). Moreover, the majority of SNPs from Non-focus and Focus were overlapped. Not only low number of SNPs were observed across all samples, but also the mutated genes were quite similar between samples (Table 3). This result once again reinforced that x-irradiation did not induce strong genomic alteration.

Mutational signatures of x-irradiated cells

Although Non-focus and Focus did not exhibit high tumour mutation burden, I speculated that there might be some signatures among low number somatic mutations. In human cancer, various types of mutational signatures were found by large-scale analyses [57, 63-66] and catalogue of somatic mutations in cancer, COSMIC, provided a spectrum of information on each mutational signature and

aetiology as well as additional mutational features [57]. Among 30 signatures from COSMIC, only one mutational signature, signature 5, was found in Non-focus and Focus samples (Figure 6). Signature 5 has been found in almost all types of cancer which may suggest that Non-focus and Focus samples possess certain properties of cancer.

Confirmation of samples using the number of SNPs

Since the focus represented malignantly transformed cells, substantial amount of genomic alteration was expected but the result turned out to be otherwise. Consequently, these gentle genomic changes led myself to focus on profound transcriptomic analysis. To make sure the samples subjected for WGS and RNA-seq were arose from the same specimen, the number of SNPs from each sequencing scheme were compared.

Firstly, the number and types of SNP detected from WGS and RNA-seq were investigated. Then, the overlapped SNPs and exclusive to WGS or RNA-seq were counted (Figure 7; Table 4). As average of 93% of SNPs were overlapped in each sample, I was able to conclude that the same samples were subjected for WGS and RNA-seq.

Characteristics of x-irradiation induced transformed cells

For transcriptome analysis, 17 samples were subjected including three Control, six Non-focus, and eight Focus (Table 1) and the summary of RNA-seq is shown in Table 5. When all samples were subjected to the principal component analysis (PCA),

a clear distinction between groups was exhibited (Figure 8A). Also, abundant number of DEGs were discovered including 2,083 up-regulated and 3,867 down-regulated genes in Foci (Figure 8B). Genes such as *Myc*, *Myct1*, *Mmp16*, *Mmp25*, *Mmp28*, *Fgd3* and *Chek1* showed elevated expression levels in Focus. It is reported that increased expression of *Myc* contributes to tumourigenesis [67, 68] and over-expressed and/ or activated *Myc* gene is frequently observed in various types of human cancers [69-75]. Also, it is known that some matrix metalloproteinases (MMPs) are involved in cancer progression [76-79]. In addition, there were a few down-regulated genes including *Tgfb2*, *Tgfb3*, *Tgfbr2*, *Olfm2* and *Adamts8*. In case of TGF-β genes, they differ in roles depend on tumour stage. TGF-β genes display down-regulation in early stage cancer and show up-regulation in advanced stage [80]. Focus cells might not represent fully developed tumour cells. However, they did undergo malignant transformation process, hence they were considered as very early stage cancer cells which may explain down-regulated TGF-β genes.

Afterwards, to estimate the cellular status using gene expression pattern, I used CountClust which illustrated the potential structure of RNA-seq expression data by clustering genes [46]. Two distinctive clusters were identified and genes that compose each cluster were subjected to the further analysis. Genes from cluster 1 were up-regulated in Control and displayed down-regulation in Non-focus then further down-regulation was observed in Focus. Conversely, cluster 2 showed the opposite trend. Genes from cluster 2 showed a trend of increment when cells went through phases from Control to Non-focus to Focus (Figure 9A). When genes from

each cluster were subjected to the pathway enrichment analysis, cluster 2 showed particular enrichment in 'DNA replication', 'cell cycle', and 'pyrimidine metabolism' (Figure 9B; Table 6). Here, the pathway enrichment results distinctly displayed that genes from cluster 2 illustrated focus-like characteristics and there was clear distinction in gene expression level between groups according to the presence of focus.

Features of Non-focus cells

To uncover which element behaves as a second factor in two-step malignant transformation, I focused on elucidating the features of Non-focus samples. Although Non-focus cells did not go through entire malignant transformation process, due to x-irradiation exposure, I was very positive that cells were damaged. The PCA using all samples revealed that Non-focus samples formed a separate cluster hence were distinguishable from un-irradiated cells (Figure 10A). Based on the differed gene expression pattern, I speculated that the hidden agenda of Little's hypothesis probably lies underneath transcriptome profile of Non-focus.

In the present study, after identifying diverged transcriptome profiles of Control and Non-focus, I examined DEGs. The volcano plot displayed 1,510 up- and 618 down-regulated DEGs (Figure 10B). Genes including *Cytl1*, *Peg3*, *Cdkn1a*, *Mmp9*, *Trp53i11*, *Il6*, and *Myc* were found to be up-regulated in Non-focus. From up-regulated DEGs, *Cytl1* and *Il6* are related to immune system [81, 82] and *Peg3* has a function in cell proliferation and p53-mediated apoptosis [83]. Also, *Mmp9* is involved in tumour development and metastasis [84] and *Myc* functions as proto-

oncogene which explains its frequent up-regulation in human cancers [69, 85]. In particular, *Cdkn1a* which encodes p21 protein and is closely controlled by p53 showed significant increment in Non-focus. As *Cdkn1a* is well-known DNA damage response gene, to sense the DNA damage caused by x-irradiation, I speculated that the expression level jumped in Non-focus [86-88]. In case of down-regulated DEGs, *Tgfb2*, *Tgfb3*, and *Tgfbr2* were discovered. As mentioned earlier, TGF-β genes are deeply involved in tumour suppression and progression depends on the tumour stage. With this approach, I was able to determine that cells from Non-focus samples were definitely in a process of undergoing transformation process and holding potential for transforming into malignant cells.

Tgfβ gene expression in early stage of malignant transformation process

To further verify the traits of Non-focus cells, I analysed 618 down-regulated genes in Non-focus. The pathway enrichment result of those genes described that processes such as 'Melanoma', 'Melanogenesis', and 'TGF-β signaling pathway' were down-regulated in Non-focus samples (Figure 11A; Table 7). Interestingly, TGF-β signaling pathway was found as one of the top 10 significantly enriched pathways which emphasized the importance of down-regulation of *Tgfβ* genes in malignant transformation process as several *Tgfβ* genes were discovered as down-regulated DEGs between Control and Non-focus as well as Non-focus and Focus (Figure 8B and Figure 10B). Thus, I postulated that *Tgfβ* genes play a crucial role in progression of malignant transformation. When the expression profile of genes that are involved in TGF-β signaling pathway was observed, numerous genes exhibited decreased

expression in Non-focus and Focus samples (Figure 11B). Next, *Tgfβ* genes that were identified as DEGs were further studied and showed gradual down-regulation from Control to Non-focus, then Non-focus to Focus samples (Figure 12). Continued decrement of *Tgfβ* genes may imply that when these genes are 'switched-off' when malignant transformation is initiated and progressed.

Elevated immune response and spatiotemporal dynamics of *Cdkn1a* in Non-focus

To further elucidate detailed characteristics of Non-focus, up-regulated DEGs were subjected to gene set enrichment analysis (GSEA). KEGG pathways and HM pathways were used for this analysis and I found that those genes were mainly enriched with immune system or immune response related pathways (Figure 13A; Table 8). The immune related genes displayed increased in expression level in Non-focus but went down in Focus, similar to that of Control or slightly further down (Figure 13B). It is known that the immune system functions in two different ways; it can participate in protection from tumourigenesis as immune surveillance or may contribute to inflammatory response which is tumour promoting environment [89, 90]. As Non-focus is positioned at early stage of transformation process, cells may fight back for the external attack. Thus, I speculated that up-regulation of immune response was due to turned-on immune surveillance system to prevent the occurrence of malignant transformation process. Then, decreased immune related genes in Focus may be caused by overloaded immune surveillance system hence interpreted as failed immune surveillance.

Furthermore, there was one gene that exhibited very peculiar expression pattern, *Cdkn1a*, P53 pathway related gene, which is already mentioned as one of the up-regulated DEGs. While four other P53 pathway related genes showed gradual up-regulation from Control to Non-focus to Focus, *Cdkn1a* showed dramatic increment in Non-focus from initial seeding then reduced in expression in other samples although it was still higher than that of Control (Figure 14). *Cdkn1a* is known as DNA damage sensing gene and a few previous researches provided evidences that *Cdkn1a* is rapidly recruited to DNA damage sites when double-strand breaks (DSBs) are induced [88, 91, 92]. Moreover, it is a regulator of cell cycle upon DNA damage, but also a regulator of DNA repair [86, 93, 94]. In addition, *Cdkn1a* is also involved in P53 pathway and genes that are engaged in this pathway showed slight increment in Non-focus when compared to Control (Figure 15A). Interestingly, further up-regulation of P53 pathway related genes were displayed in Focus when compared to Non-focus (Figure 15B). P53 pathway is one of the major pathways that plays a role in tumourigenesis by regulating cell cycle and activating DNA repair process [95]. It is up-regulated upon external stress or DNA damages to prevent tumourigenesis. In the present study, upon x-irradiation, P53 pathway was activated. When the effect was accumulated in cells, which lead to focus generation, P53 pathway was further raised. Here, I believe that slight increment of P53 pathway in Non-focus and dramatic increase in Focus may once again suggest that Non-focus cells were one step behind from developing into transformants and Focus represented *in vitro* version of tumour.

Down-regulated immune system and histone genes in transformed cells

Weighted correlation network analysis (WGCNA) was used to investigate the transcriptomic difference between each sample group according to the presence of focus. I identified 10 modules, five of which were statistically significant (Figure 16A). Turquoise (r = -0.78 and p = 2E-04) and pink (r = -0.54 and p = 0.02) module showed a trend of decrement with presence of focus and brown (r = 0.74 and p = 7E-04), blue (r = 0.51 and p = 0.04), and yellow (r = 0.5 and p = 0.04) exhibited otherwise. When the transcriptome profile of all genes from each significant module were illustrated clear up- and down-regulation of each module were observed (Figure 16B). Interestingly, genes from pink module displayed intriguing transcriptomic profile. There was a sudden up-regulation in Non-focus, then showed steep down-regulation in Focus whereas turquoise module showed gradual decrease from Control to Non-focus then to Focus samples. When down-regulated modules were subjected to the pathway enrichment analysis, turquoise module was enriched with 'pathways in cancer', 'regulation of actin cytoskeleton', and 'focal adhesion'. In case of pink module, pathways such as 'systemic lupus erythematosus', 'viral myocarditis' and 'cytokine-cytokine receptor interaction' were found to be enriched (Figure 17; Table 9). Intriguingly, there were five histone genes in pink module (5/183), *Hist1h4i*, *Hist1h3d, Hist1h4h*, *Hist1h2bf*, and *Hist1h4k* which were enriched with 'systemic lupus erythematosus'. When DEG analysis was performed with all samples, 24 histone genes were found as down-regulated genes in Focus including above five histone genes from pink module (Figure 18). It is known that declined immune system may contribute to establish tumour-promoting environment by diminishing

immune surveillance [89]. Also, down-regulation of histone genes upon DNA damage was also previously reported and this result reflected their conclusion [96]. Therefore, down-regulated histone gene expression may be considered as one of the characteristics of transformed cells.

Defected DNA repair pathways and cell cycle regulation in Focus

Various biological processes can get deregulated due to external stimuli including DNA repair process and cell cycle regulation. X-irradiation is a type of external stimuli which can introduce DNA damage in cells. Here, brown module exhibited up-regulated DNA repair process including 'homologous recombination', 'mismatch repair', and 'base excision repair' as well as 'cell cycle' and 'DNA replication' (Figure 19; Table 10). Pathways that were mainly enriched with blue module were mitochondrial function related pathways such as 'oxidative phosphorylation', 'Parkinson's disease' and 'proteasome' (Figure 19; Table 10). For yellow module, 'WNT signaling pathway' and 'Ubiquitin mediated proteolysis' were found (Figure 19; Table 10).

Focus formation may indicate that cells are experiencing malignant transformation process, yet it does not mean that focus cells are fully developed tumour cells. Therefore, as the transformation process is still ongoing, the characteristics of Focus cells may not be consistent, hence, both tumour-promoting and -preventing characteristics can co-exist at this stage. In this study, elevated apoptotic genes and DNA repair processes were observed in Focus samples. As Focus are considered as transformants, damaged DNA repair system and

dysregulated cell cycle process were may be expected. In the present study, I focused on genes that are deeply involved in those processes. *Atm* and *Atr* are well-known DNA damage response genes and DNA damage response signaling pathway is orchestrated by above two genes [97]. They are found in yellow module and greatly involved in broad spectrum of biological processes in order to maintain genomic integrity including apoptotic process [98]. Upon x-irradiation, *Atm* sensed DNA damage, it directly activated *Trp53*, then activation of various pro-apoptotic genes was followed (Figure 20). When pro-apoptotic genes were further analysed, continued elevation of those genes from Non-focus to Focus were observed (Figure 21A and 21B). In particular, *Bak* and *Bax* were found in blue and *Apaf1* was found in yellow module, which were both up-regulated modules in Focus. Based on these findings, I speculated that apoptotic process is up-regulated as a compensatory mechanism to eliminate damages cells occurred during transformation process.

Interestingly, *Atm* and *Atr* showed down-regulation in Non-focus then displayed up-regulation in Focus (Figure 22). As DNA repair and cell cycle regulation processes are affected by *Atm* and *Atr*, it is important to observe gene expression shift in genes that are related in those processes. *Brca1*, *Brca2*, and *Chek1*, that were all included in brown module, which also is an up-regulated module in Focus, were subjected to the analysis and they showed similar expression pattern as that of *Atm* and *Atr*. They all showed slight down-regulation in Non-focus then great up-regulation in Focus (Figure 22). Here, a schematic illustration of two types of DNA repair pathways described the inter-connected relationship of mentioned genes upon x-irradiation in Non-focus and Focus (Figure 23A and Figure 23B). When Non-

focus and Focus were compared, DNA repair pathway related genes were more up-regulated in focus. It is reported that when error-free repair pathway is overloaded, then error-prone repair is triggered as damaged DNA cannot be fixed with error-free way [13-15]. In this study, Focus samples exhibited saturated homologous recombination and as a result, increased non-homologous end-joining process was observed. As a response to damaged DNA and also as a compensatory mechanism, DNA repair process was up-regulated at the first place. Then, when the damage was too severe, error-free pathway became saturated, the cells had to choose turning on the error-prone repair pathway as a last resort. Thus, error was never completely fixed no matter how up-regulated the repair pathway was, hence focus was rigidly formed.

In addition, there are *Xrcc* genes that show significant involvement in DNA repair processes [99]. Most of these genes are the components of different types of repair pathways such as homologous recombination (HR) and non-homologous end-joining (NHEJ), hence comprehending the role of *Xrcc* genes may contribute to better understanding of human diseases including cancer. Among several *Xrcc* genes, *Xrcc1*, *Xrcc2*, *Xrcc5*, and *Xrcc6* were increased in expression and enriched with up-regulated modules in Focus (Figure 24). *Xrcc1* and *Xrcc6* were found in blue module and *Xrcc2* and *Xrcc5* were found in brown module. *Xrcc1* exhibited an implication in general coordination of DNA base repair [100, 101] and *Xrcc2* displayed *RAD51*-like characteristics and involved in HR process [102-105]. Also, when *Xrcc2* was defected in mouse model, increased apoptosis was observed [106]. Furthermore, *Xrcc5* and *Xrcc6* were engaged in NHEJ process [107, 108].

Stemness of transformed cells

Understanding how stem cells work will help to comprehend how diseases occur, how healthy cells are generated to substitute diseased cells or even testing new drugs. Although stem cells play an important role in tissue repair, there are also other studies that described the function of stem cells in malignant transformation and tumour initiation [109-112]. To observe whether focus cells hold stem cell-like characteristics, several chemical and genetic perturbations (CGP) gene sets from GSEA were used. Ramalho *et al.* discovered 195 up-regulated and 69 down-regulated genes in embryonic, neural, and hematopoietic stem cells [113]. In my result, those genes were up-regulated and down-regulated respectively in Focus (Figure 25). Also, from CGP, Bhattacharya *et al.* represented 85 genes that displayed stemness signature with their up-regulation in six human embryonic stem cell (ESC) lines [114]. Moreover, Wong *et al.* found 328 genes that were up-regulated in mouse ESC that were overlapped with human ESC [115]. When those CGP gene sets were subjected to the analysis, Focus cells exhibited up-regulation (Figure 25). These findings suggested that transformed focus cells may have stem cell-like characteristics.

Table 1. The summary of sample information.

Sample ID	Dish of origin	Seeding	Focus status
X-ray 1	A	1st	No
X-ray 2	A	1st	Focus
X-ray 3	A	1st	Focus
X-ray 4	B	1st	No
X-ray 5	B	1st	Focus
X-ray 6	B	1st	Focus
X-ray 7	A	2nd	No
X-ray 8	A	2nd	Focus
X-ray 9	A	2nd	No
X-ray 10	A	2nd	Focus
X-ray 11	B	2nd	No
X-ray 12	B	2nd	Focus
X-ray 13	B	2nd	No
X-ray 14	B	2nd	Focus
37_CON	Control	Control	Control
X-ray 16	Control	Control	Control
X-ray 17	Control	Control	Control

Table 2. The summary of 14 WGS samples.

Sample ID	Mean depth (X)	Total read	Total mapped read	Paired in sequencing	Read1	Read2
X-ray 1	31.2	633096169	633096169	633096169	318698142	314398027
X-ray 2	34.5	723945761	723945761	723945761	364275520	359670241
X-ray 3	31.7	647992755	647992755	647992755	326316048	321676707
X-ray 4	35.5	741950177	741950177	741950177	373358908	368591269
X-ray 5	41.6	869578759	869578759	869578759	438092842	431485917
X-ray 6	45	937592428	937592428	937592428	471792416	465800012
X-ray 7	37.4	764499498	764499498	764499498	384509782	379989716
X-ray 8	35	740063233	740063233	740063233	372810814	367252419
X-ray 10	39.3	803911713	803911713	803911713	404080089	399831624
X-ray 11	33.8	690418134	690418134	690418134	346452125	343966009
X-ray 12	35.5	717960298	717960298	717960298	361182698	356777600
X-ray 14	34.3	694497621	694497621	694497621	348959052	345538569
X-ray 16	34.4	697239152	697239152	697239152	349620822	347618330
X-ray 17	29.7	592436434	592436434	592436434	297468622	294967812

Table 3. Somatic mutations found in all samples.

Chr	Position	Ref	Alt	Func. refGene	Gene. refGene	ExonicFunc. refGene	AAChange. refGene	X-ray 1	X-ray 4	X-ray 7	X-ray 11	X-ray 2	X-ray 3	X-ray 5	X-ray 6	X-ray 8	X-ray 10	X-ray 12	X-ray 14
chr1	16103568	C	A	exonic	*Rpl7*	missense SNV	R44L	./.	./.	0/1	0/1	./.	0/1	0/1	./.	./.	./.	./.	./.
chr1	85212677	G	T	exonic	*Gm7609*	missense SNV	V10L	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr1	85643260	A	T	exonic	*Sp140*	stopgain SNV	K474X	./.	./.	./.	./.	./.	./.	0/1	./.	./.	0/1	./.	./.
chr1	85643312	G	T	exonic	*Sp140*	missense SNV	R491M	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.
chr1	85643339	C	T	exonic	*Sp140*	missense SNV	T500M	0/1	./.	./.	./.	./.	./.	0/1	./.	./.	0/1	./.	./.
chr1	85643343	C	A	exonic	*Sp140*	missense SNV	F501L	0/1	./.	./.	./.	./.	./.	0/1	0/1	./.	0/1	./.	./.
chr2	89144852	G	T	exonic	*Olfr1223*	missense SNV	P57H	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr2	111368761	T	G	exonic	*Olfr1283*	missense SNV	L43R	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr2	164256266	C	T	exonic	*Svs3b*	missense SNV	S45N	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.
chr2	166081688	G	A	exonic	*Sulf2*	missense SNV	A537V	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.
chr2	176721008	G	A	exonic	*Gm14305*	missense SNV	R231Q	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr2	176721428	G	A	exonic	*Gm14305*	missense SNV	R371Q	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr2	177832102	C	T	exonic	*Gm14325*	missense SNV	G396S	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr3	53477969	C	T	exonic	*Proser1*	missense SNV	A424V	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.

chr3	93215431	C	A	exonic	*Flg2*	missense SNV	S1636Y	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr3	93215770	G	T	exonic	*Flg2*	missense SNV	G1749V	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr3	93397577	G	A	exonic	*Rptn*	missense SNV	R739Q	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.
chr3	93446630	T	G	exonic	*Tchh*	missense SNV	F1126V	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr3	93447158	T	G	exonic	*Tchh*	missense SNV	F1302V	./.	./.	0/1	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr4	81365987	T	G	exonic	*Mpdz*	missense SNV	E519D	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr4	126043633	G	A	exonic	*Csf3r*	missense SNV	R662H	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr4	137428531	G	A	exonic	*Cela3b*	missense SNV	H22Y	./.	./.	./.	./.	./.	./.	./.	0/1	./.	0/1	./.	./.
chr4	145622307	T	A	exonic	*Gm13212*	missense SNV	Y105N	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr4	146124202	A	G	exonic	*Gm13051*	missense SNV	D68G	./.	./.	./.	./.	./.	./.	0/1	./.	./.	0/1	./.	./.
chr4	146124571	A	C	exonic	*Gm13051*	missense SNV	K191T	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.
chr4	146125350	C	G	exonic	*Gm13051*	missense SNV	L451V	0/1	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr4	146125423	A	T	exonic	*Gm13051*	missense SNV	E475V	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr4	146125702	G	T	exonic	*Gm13051*	missense SNV	R568I	./.	./.	./.	./.	0/1	./.	./.	./.	./.	0/1	./.	./.
chr4	146195532	G	A	exonic	*Zfp600*	missense SNV	D257N	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr4	146195808	C	T	exonic	*Zfp600*	missense SNV	H349Y	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.

chr4	146196088	C	A	exonic	*Zfp600*	missense SNV	T442K	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.
chr4	147057808	A	T	exonic	*Rex2*	missense SNV	Y251F	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr4	147058116	C	T	exonic	*Rex2*	missense SNV	R354C	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr4	147058728	C	T	exonic	*Rex2*	missense SNV	P558S	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	0/1
chr4	147674949	G	C	exonic	*Zfp534*	missense SNV	S421C	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr4	147675066	C	T	exonic	*Zfp534*	missense SNV	C382Y	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr4	147675508	T	A	exonic	*Zfp534*	missense SNV	T235S	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr4	147675617	T	G	exonic	*Zfp534*	missense SNV	K198N	0/1	./.	./.	./.	0/1	./.	./.	0/1	./.	./.	./.	./.
chr4	147675639	T	G	exonic	*Zfp534*	missense SNV	K191T	./.	./.	./.	./.	0/1	./.	0/1	./.	./.	./.	./.	./.
chr4	147755911	T	G	exonic	*Gm13157*	missense SNV	E161A	0/1	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr4	147757804	C	T	exonic	*Gm13157*	missense SNV	A19T	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.
chr5	14934158	C	T	exonic	*Speer4e*	missense SNV	E215K	0/1	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.
chr5	15621800	A	G	exonic	*Speer4d*	missense SNV	H101R	./.	./.	./.	0/1	0/1	0/1	./.	0/1	./.	./.	./.	0/1
chr5	26086441	C	A	exonic	*Gm10471*	missense SNV	K122N	./.	0/1	0/1	0/1	./.	./.	0/1	0/1	./.	./.	0/1	0/1
chr5	26086487	A	T	exonic	*Gm10471*	missense SNV	F107Y	./.	./.	./.	./.	./.	./.	0/1	0/1	./.	./.	./.	./.
chr5	26105203	A	G	exonic	*5031410I06Rik*	missense SNV	L30P	./.	./.	./.	./.	./.	./.	0/1	0/1	./.	0/1	0/1	./.

chr5	33664389	A	C	exonic	*Tacc3*	missense SNV	E164D	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.
chr5	93182981	A	G	exonic	*Ccni*	missense SNV	M377T	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr5	93183011	G	T	exonic	*Ccni*	missense SNV	P367H	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr5	94383818	A	G	exonic	*AA792892*	missense SNV	N187S	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr5	109086448	G	A	exonic	*Vmn2r12*	missense SNV	P633S	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr5	112839687	G	T	exonic	*Myo18b*	missense SNV	P1037H	./.	./.	./.	./.	./.	./.	0/1	./.	./.	0/1	./.	./.
chr5	113819688	G	T	exonic	*Selplg*	missense SNV	Q186K	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr5	121208858	T	C	exonic	*Rpl6*	missense SNV	V272A	./.	0/1	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr5	137520039	G	T	exonic	*Gigyf1*	missense SNV	R165L	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr5	146556209	G	A	exonic	*Gm3402*	missense SNV	A16T	0/1	./.	0/1	./.	./.	0/1	0/1	0/1	0/1	0/1	./.	./.
chr6	127949827	G	T	exonic	*Tspan11*	missense SNV	R246L	./.	0/1	0/1	./.	0/1	0/1	0/1	0/1	0/1	./.	0/1	0/1
chr6	128620279	C	G	exonic	*Klrb1a*	missense SNV	V59L	./.	./.	0/1	./.	./.	./.	0/1	./.	./.	./.	./.	0/1
chr6	132361427	C	T	exonic	*Gm8882*	missense SNV	S276N	./.	./.	./.	./.	./.	./.	0/1	0/1	./.	0/1	./.	./.
chr7	4919583	A	C	exonic	*Zfp628*	missense SNV	Q268P	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.
chr7	18426174	T	C	exonic	*Psg28*	missense SNV	M366V	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr7	18426192	C	T	exonic	*Psg28*	missense SNV	V360I	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1

chr7	18426377	C	T	exonic	*Psg28*	missense SNV	G298D	./.	0/1	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr7	18478342	T	C	exonic	*Psg26*	missense SNV	I363V	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr7	18652398	A	T	exonic	*Psg21*	missense SNV	F221Y	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr7	38519652	T	G	exonic	*Gm5591*	missense SNV	N599T	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.
chr7	41044040	T	C	exonic	*Gm4884*	missense SNV	S478P	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr7	41286343	A	G	exonic	*Gm5592*	missense SNV	T90A	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr7	41288226	G	C	exonic	*Gm5592*	missense SNV	E311Q	./.	./.	0/1	0/1	./.	0/1	./.	./.	./.	0/1	./.	./.
chr7	43226166	A	G	exonic	*EU599041*	missense SNV	I216V	./.	./.	0/1	./.	./.	0/1	0/1	./.	./.	./.	./.	./.
chr7	44251482	C	A	exonic	*2410002F23Rik*	stopgain SNV	Y270X	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr7	44251493	C	T	exonic	*2410002F23Rik*	missense SNV	A274V	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr7	109048037	A	T	exonic	*Ric3*	missense SNV	I193N	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr7	141859161	C	A	exonic	*Muc5b*	missense SNV	P1948Q	0/1	./.	./.	./.	./.	./.	0/1	./.	./.	0/1	./.	./.
chr8	3459381	A	C	exonic	*Pex11g*	missense SNV	V182G	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr8	19732819	A	G	exonic	*4930467E23Rik*	missense SNV	K124R	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr8	20357039	G	A	exonic	*Gm15319*	missense SNV	P224S	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr8	24962957	C	T	exonic	*Adam9*	missense SNV	G778D	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.

chr8	24962958	C	T	exonic	*Adam9*	missense SNV	G778S	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr8	27345850	T	A	exonic	*Tex24*	missense SNV	L330M	./.	./.	./.	./.	./.	./.	0/1	./.	./.	0/1	0/1	./.
chr8	40794988	C	T	exonic	*Adam20*	missense SNV	P45L	./.	./.	0/1	./.	0/1	./.	0/1	./.	./.	0/1	./.	./.
chr8	94521411	G	A	exonic	*Nlrc5*	missense SNV	R1669H	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr9	89090621	T	A	exonic	*Trim43b*	missense SNV	K160I	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.
chr9	109286487	T	C	exonic	*Fbxw14*	missense SNV	N72D	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr9	120018428	G	A	exonic	*Xirp1*	missense SNV	S463L	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr10	100340907	C	G	exonic	*Gm4303*	missense SNV	L18V	0/1	./.	./.	./.	./.	./.	0/1	0/1	0/1	0/1	0/1	0/1
chr10	100340932	G	A	exonic	*Gm4303*	missense SNV	R26H	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr11	5032491	G	A	exonic	*Ap1b1*	missense SNV	A624T	./.	./.	0/1	./.	0/1	0/1	./.	0/1	0/1	./.	./.	./.
chr11	58502494	G	A	exonic	*Olfr331*	missense SNV	H27Y	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.
chr11	58684277	C	T	exonic	*Olfr320*	missense SNV	L135F	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr11	58758255	C	T	exonic	*Olfr316*	missense SNV	H197Y	./.	./.	0/1	./.	0/1	./.	./.	0/1	./.	./.	./.	0/1
chr11	99785583	A	G	exonic	*Krtap4-9*	missense SNV	Q110R	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.
chr11	114893406	C	T	exonic	*Cd300a*	missense SNV	A80V	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.
chr11	114893412	G	A	exonic	*Cd300a*	missense SNV	S82N	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.

chr11	116347440	A	G	exonic	*Rnf157*	missense SNV	S533P	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr11	120347084	C	G	exonic	*Actg1*	missense SNV	R210P	./.	./.	0/1	0/1	./.	./.	./.	./.	./.	./.	./.	./.
chr11	120347130	C	T	exonic	*Actg1*	missense SNV	E195K	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr11	120347154	C	T	exonic	*Actg1*	missense SNV	D187N	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr12	31074539	G	A	exonic	*Fam110c*	missense SNV	A167T	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr12	86983623	A	G	exonic	*Zdhhc22*	missense SNV	F184L	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr12	111730089	C	A	exonic	*Apopt1*	missense SNV	F152L	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.
chr13	66122638	T	A	exonic	*Gm10324*	missense SNV	H539Q	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.
chr13	100162007	T	C	exonic	*Naip2*	missense SNV	N507S	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.
chr13	119949600	C	A	exonic	*B020031M17Rik*	missense SNV	G157C	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	0/1	0/1
chr14	42794859	G	C	exonic	*Gm10377*	missense SNV	N113K	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.
chr14	50515832	G	T	exonic	*Olfr742*	missense SNV	M209I	./.	./.	./.	0/1	0/1	0/1	./.	./.	0/1	./.	./.	0/1
chr14	51101527	C	G	exonic	*Ang*	missense SNV	P42A	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chr14	54664562	G	T	exonic	*Acin1*	missense SNV	S591Y	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.
chr14	79453128	C	A	exonic	*Kbtbd6*	missense SNV	P358H	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.
chr15	78348350	G	A	exonic	*Csf2rb*	missense SNV	S619N	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.

chr15	89424789	A	C	exonic	*Cpt1b*	missense SNV	V70G	./.	0/1	./.	./.	0/1	./.	0/1	./.	./.	./.	0/1	0/1
chr15	91908701	G	T	exonic	*Muc19*	missense SNV	S2076I	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr15	101947704	A	C	exonic	*Krt78*	missense SNV	I557M	./.	./.	./.	0/1	./.	./.	./.	./.	0/1	./.	./.	./.
chr16	32753750	G	T	exonic	*Muc4*	missense SNV	V1209F	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.	./.
chr17	15544509	T	G	exonic	*Prdm9*	missense SNV	K670Q	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr17	20268350	C	T	exonic	*Vmn2r106*	missense SNV	A596T	./.	./.	./.	0/1	./.	./.	./.	./.	./.	./.	./.	0/1
chr17	35323300	T	G	exonic	*H2-Q1*	missense SNV	Y283D	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr17	35554981	T	A	exonic	*Cdsn*	missense SNV	S136T	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chr17	35619613	A	G	exonic	*Gm9573*	missense SNV	L1227P	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1
chr17	56104806	A	C	exonic	*Plin4*	missense SNV	S742A	./.	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.
chr17	94876751	G	A	exonic	*Gm20939*	missense SNV	G277S	./.	./.	./.	./.	./.	./.	./.	./.	0/1	./.	0/1	./.
chr18	58056266	C	A	exonic	*Fbn2*	missense SNV	R1596L	./.	./.	0/1	./.	0/1	./.	0/1	0/1	0/1	0/1	0/1	0/1
chr19	60762120	A	C	exonic	*Eif3a*	missense SNV	N1332K	./.	0/1	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.
chrX	61184396	T	C	exonic	*Cdr1*	missense SNV	Q388R	./.	./.	./.	./.	./.	./.	./.	0/1	./.	./.	./.	./.
chrX	73290421	G	T	exonic	*Xlr5c*	missense SNV	A32E	./.	./.	./.	./.	0/1	./.	./.	./.	./.	0/1	./.	./.

Table 4. The number of SNPs found in WGS and RNA-seq.

	RNA-seq only	Common		WGS only
		RNA-seq	WGS	
Average (%)	6.1	93.9	42.9	57.2

Table 5. The average RNA-seq summary of 17 samples.

Sample ID	Total reads	Uniquely mapped reads	Uniquely mapped reads (%)
X-ray 1	57976963	52616491	90.75
X-ray 2	65975292	59598866	90.34
X-ray 3	77331102	70429872	91.08
X-ray 4	88511140	80875924	91.37
X-ray 5	76806638	69356984	90.30
X-ray 6	77945767	70812724	90.85
X-ray 7	75905005	68752643	90.58
X-ray 8	59284103	52570998	88.68
X-ray 9	75382149	67846665	90.00
X-ray 10	85873980	77317134	90.04
X-ray 11	76429057	68156009	89.18
X-ray 12	91091763	80311451	88.17
X-ray 13	70897805	63026807	88.90
X-ray 14	76403307	67236244	88.00
37_CON	74546548	68668509	92.11
X-ray 16	80547809	73703094	91.50
X-ray 17	66579474	60544875	90.94

Table 6. The top 10 significantly enriched KEGG pathways in Cluster 1 and Cluster 2.

Cluster	Pathway	*p*-value	*q*-value
Cluster 1	Pathways in cancer	1.30E-20	2.41E-18
	Focal adhesion	1.42E-18	1.32E-16
	Lysosome	3.41E-16	2.12E-14
	ECM receptor interaction	4.33E-15	2.01E-13
	WNT signaling pathway	3.86E-12	1.43E-10
	Melanoma	9.58E-12	2.97E-10
	Prostate cancer	2.86E-10	7.60E-09
	Regulation of actin cytoskeleton	9.11E-08	2.12E-06
	Calcium signaling pathway	1.32E-07	2.43E-06
	Pancreatic cancer	1.44E-07	2.43E-06
Cluster 2	Ribosome	1.42E-40	2.65E-38
	Cell cycle	7.31E-35	6.80E-33
	Focal adhesion	1.36E-22	8.42E-21
	Oocyte meiosis	7.68E-20	3.57E-18
	Regulation of actin cytoskeleton	1.51E-19	5.62E-18
	Dna replication	3.67E-17	1.14E-15
	Aminoacyl trna biosynthesis	3.84E-16	1.02E-14
	Pyrimidine metabolism	2.94E-13	6.84E-12
	Tight junction	8.37E-13	1.73E-11
	ECM receptor interaction	3.41E-12	6.34E-11

Table 7. The top 10 significantly enriched KEGG pathways using down-regulated DEGs in Non-focus.

Pathway	*p*-value	*q*-value
Pathways in cancer	1.95E-12	3.63E-10
Cytokine-cytokine receptor interaction	2.77E-10	2.58E-08
ABC transporters	1.04E-07	6.47E-06
Melanoma	2.96E-06	1.02E-04
Drug metabolism cytochrome p450	3.25E-06	1.02E-04
Melanogenesis	3.30E-06	1.02E-04
WNT signaling pathway	7.80E-06	2.07E-04
TGF-B signaling	1.07E-05	2.48E-04
Focal adhesion	1.21E-05	2.50E-04
Calcium signaling pathway	2.90E-05	5.35E-04

Table 8. The top 5 significantly enriched KEGG pathways using up-regulated DEGs in Non-focus.

Type	Pathway	*p*-value	*q*-value
KEGG	Cytokine-cytokine receptor interaction	4.12E-06	7.66E-04
	NOD-like receptor signaling pathway	1.30E-04	1.05E-02
	Pathways in cancer	1.70E-04	1.05E-02
	Focal adhesion	2.90E-04	1.35E-02
	Ether lipid metabolism	1.24E-03	4.61E-02
Hallmark	Interferon γ response	3.14E-11	1.57E-09
	Inflammatory response	7.75E-07	1.94E-05
	IL6-JAK-STAT3 signaling	1.12E-05	1.86E-04
	Hypoxia	1.64E-05	2.05E-04
	Interferon a response	4.89E-04	2.45E-03

Table 9. The top 10 significant KEGG pathways found in down-regulated modules in WGCNA.

Module colour	Pathway	*p*-value	*q*-value
Turquoise	Pathways in cancer	3.56E-24	6.61E-22
	Regulation of actin cytoskeleton	1.06E-22	9.90E-21
	Focal adhesion	3.02E-21	1.87E-19
	MAPK signaling pathway	2.25E-15	1.05E-13
	FC g R mediated phagocytosis	2.17E-13	8.07E-12
	Melanoma	1.27E-12	3.92E-11
	Neurotrophin signaling pathway	1.81E-12	4.82E-11
	Lysosome	3.12E-12	7.25E-11
	Dilated cardiomyopathy	1.55E-11	3.20E-10
	Leukocyte transendothelial migration	4.65E-11	8.64E-10
Pink	Systemic lupus erythematosus	3.57E-07	6.63E-05
	Calcium signaling pathway	2.74E-04	2.02E-02
	Melanogenesis	3.26E-04	2.02E-02
	Basal cell carcinoma	7.30E-04	2.86E-02
	Hedgehog signaling pathway	7.70E-04	2.86E-02
	Leishmania infection	1.60E-03	3.49E-02
	Viral myocarditis	1.66E-03	3.49E-02
	Cytokine-cytokine receptor interaction	1.69E-03	3.49E-02
	MAPK signaling pathway	1.69E-03	3.49E-02

Table 10. The top 10 significant KEGG pathways found in up-regulated modules in WGCNA.

Module colour	Pathway	*p*-value	*q*-value
Brown	Cell cycle	4.91E-24	9.13E-22
	Homologous recombination	4.07E-14	3.79E-12
	DNA replication	2.09E-12	1.29E-10
	Oocyte meiosis	6.05E-10	2.81E-08
	Mismatch repair	2.36E-09	8.79E-08
	Purine metabolism	6.36E-07	1.85E-05
	Regulation of actin cytoskeleton	6.96E-07	1.85E-05
	Pyrimidine metabolism	1.15E-06	2.66E-05
	Base excision repair	2.04E-06	4.22E-05
	Progesterone mediated oocyte maturation	1.08E-05	2.01E-04
Blue	Huntington's disease	2.57E-64	4.79E-62
	Ribosome	1.23E-60	1.14E-58
	Oxidative phosphorylation	1.59E-49	9.87E-48
	Parkinson's disease	8.81E-49	4.10E-47
	Alzheimer's disease	2.57E-46	9.56E-45
	Splicesome	2.18E-35	6.75E-34
	Proteasome	1.19E-24	3.16E-23
	Purine metabolism	1.82E-20	4.23E-19
	Pyrimidine metabolism	2.82E-18	5.83E-17
	DNA replication	2.05E-17	3.81E-16
Yellow	WNT signaling pathway	1.99E-12	2.48E-10
	Ubiquitin mediated proteolysis	2.66E-12	2.48E-10
	Adherens junction	7.20E-10	4.47E-08
	Oocyte meiosis	3.53E-09	1.64E-07
	Focal adhesion	1.82E-08	6.38E-07
	Endocytosis	2.06E-08	6.38E-07
	Regulation of actin cytoskeleton	6.07E-08	1.60E-06
	Pathways in cancer	6.89E-08	1.60E-06
	Insulin signaling pathway	3.12E-07	6.44E-06
	Cell cycle	7.77E-07	1.45E-05

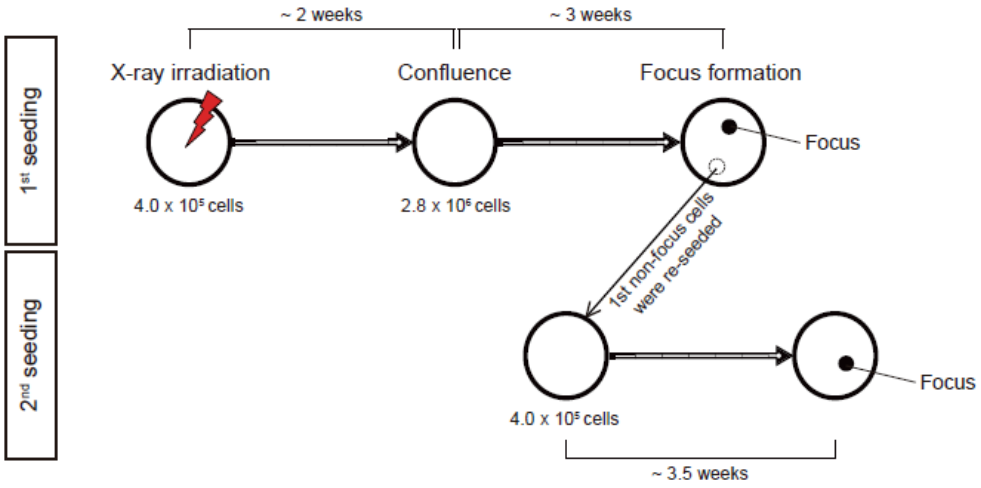

Figure 1. Experimental scheme of malignant transformation in C3H10T1/2 cells. The number of cells initialed seeded was approximately 4.0 x 10^5 cells per 100mm dish. The cells were irradiated with 400/ 600/ 800rads of x-irradiation then cultured until confluency. Upon focus formation, non-focus cells from focus generated dish were isolated then re-seeded to another plate. After similar period of time, focus was again generated.

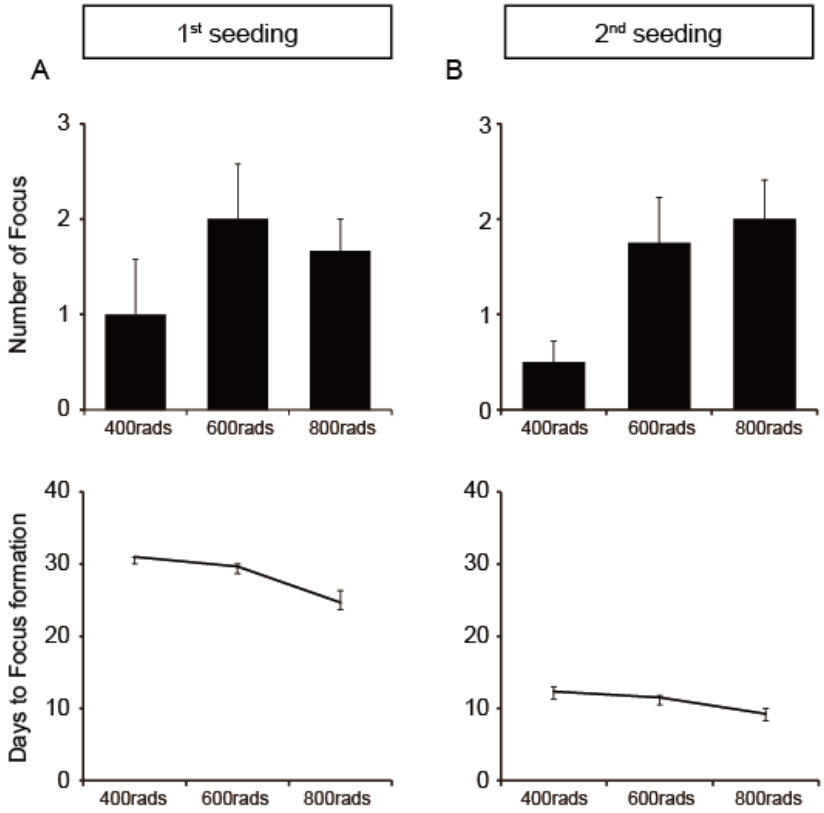

Figure 2. The number of focus generated and time taken for focus generation.
(A) Number of focus generated and time taken for focus generation from 1st seeding.
(B) Number of focus generated and time taken for focus generation from 2nd seeding.

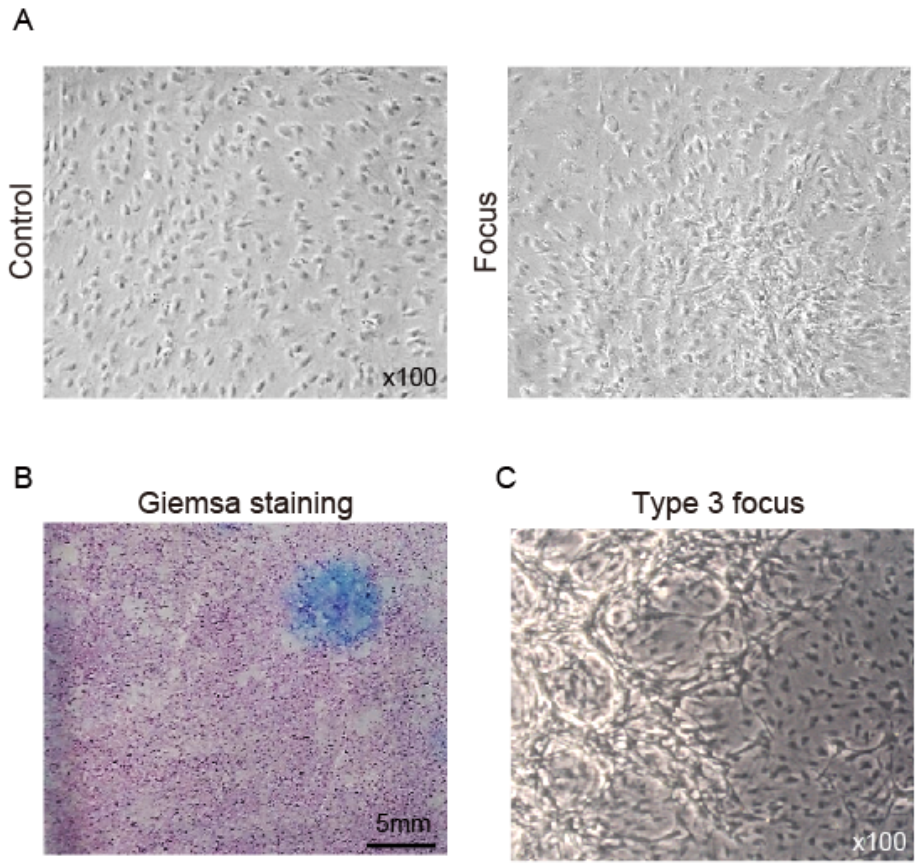

Figure 3. The cell images of Control cells and Focus cells. (A) Morphological difference was observed in focus cells. (B) Giemsa staining confirmed focus generation. Blue stain represented focus. (C) Types 3 focus was mainly found in this study.

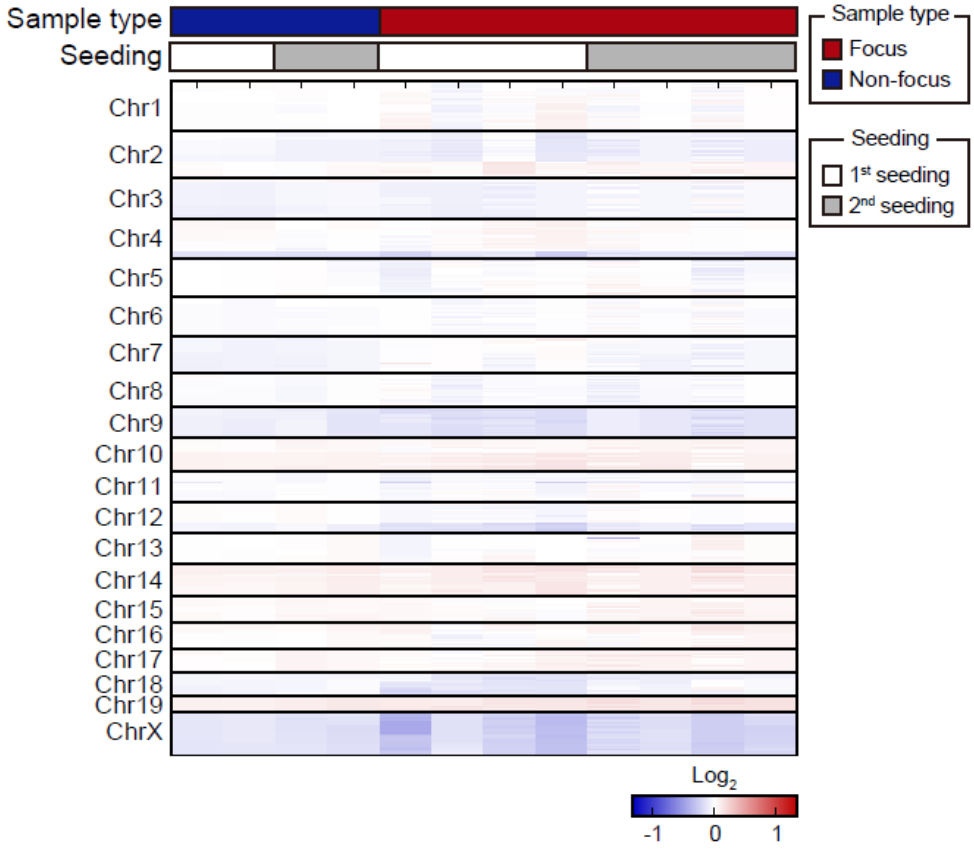

Figure 4. Somatic copy number variation in Non-focus and Focus samples. The alteration pattern between two groups were quite similar to each other.

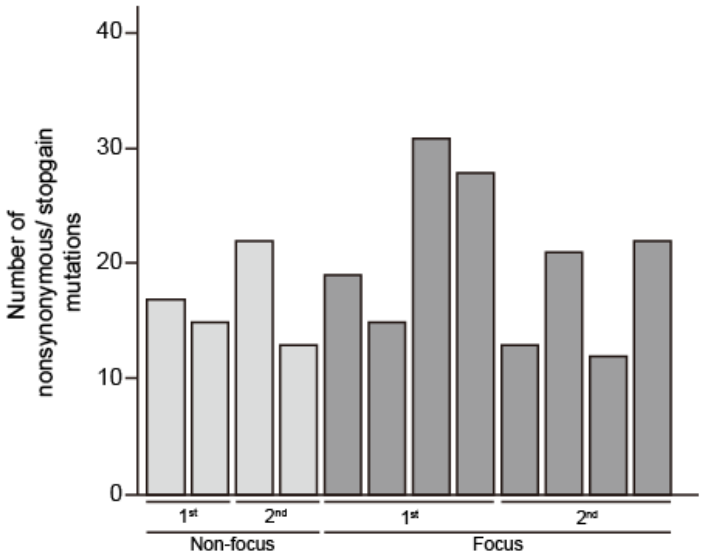

Figure 5. Tumour mutation burden of Non-focus and Focus samples. The number of somatic mutations were very low across all samples.

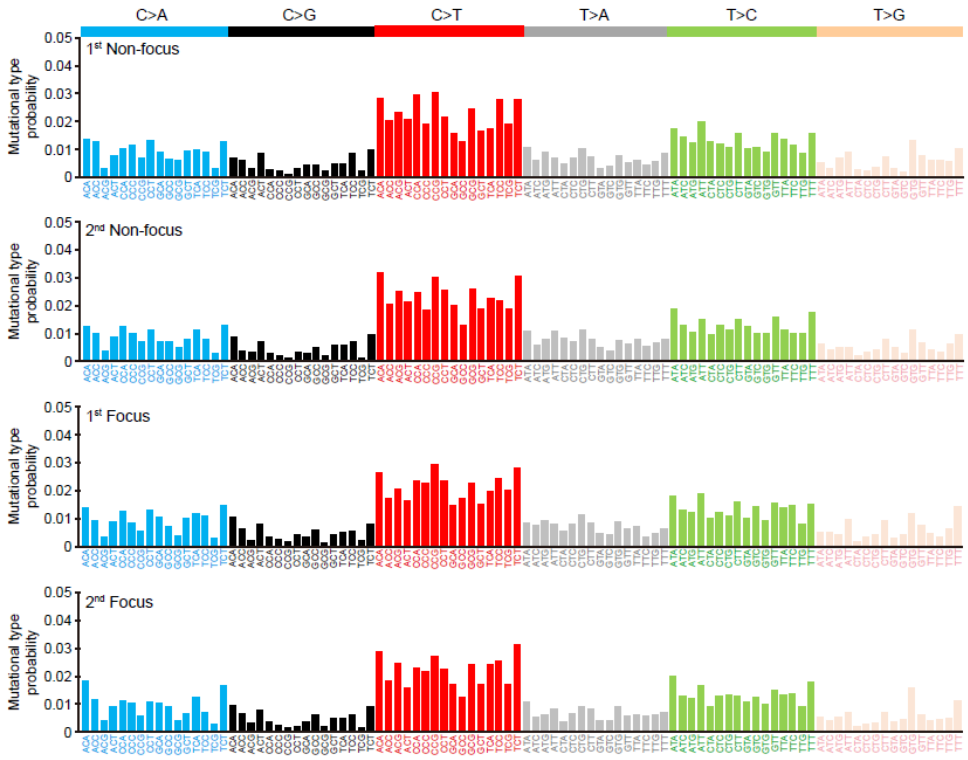

Figure 6. Signatures of mutational process in Non-focus and Focus samples. Both Non-focus and Focus samples demonstrated Signature 5 from COSMIC.

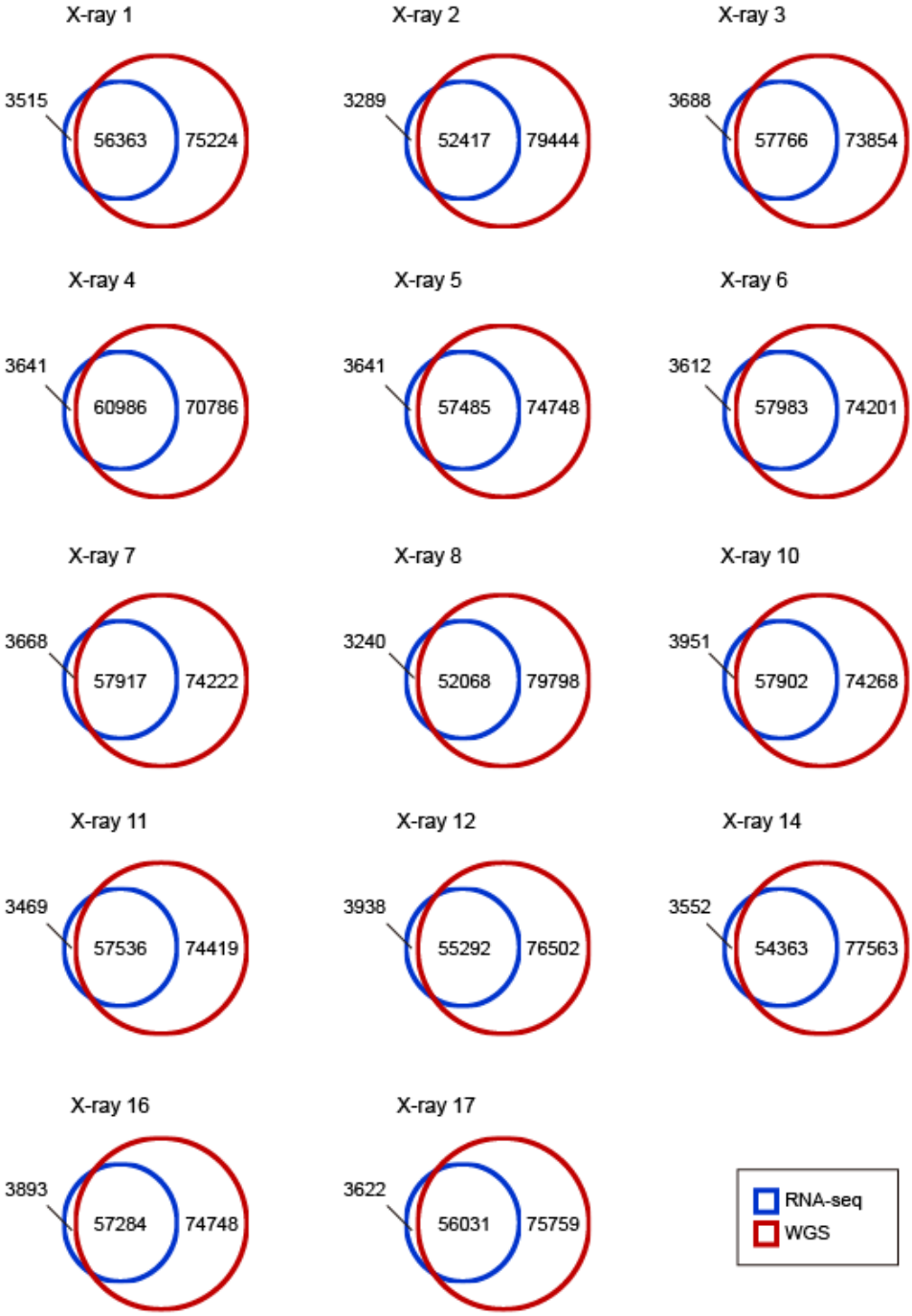

Figure 7. The number of SNPs from WGS and RNA-seq data. To confirm whether the subjected samples for WGS and RNA-seq were identical, the number of commonly found SNPs were identified.

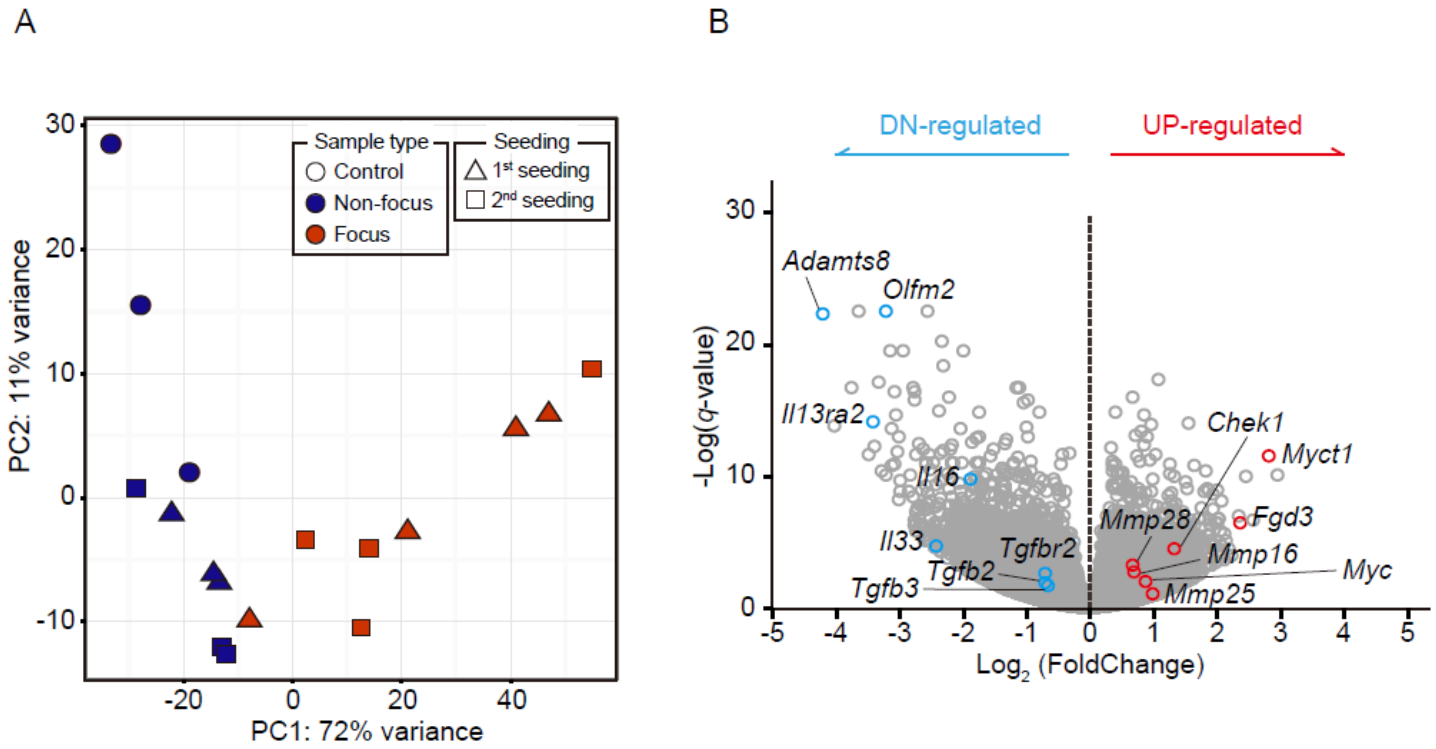

Figure 8. The transcriptomic profile of all samples. (A) Principal component analysis. The focus status divided the samples into two groups. (B) Volcano plot illustrated the differentially expressed genes according to the focus status.

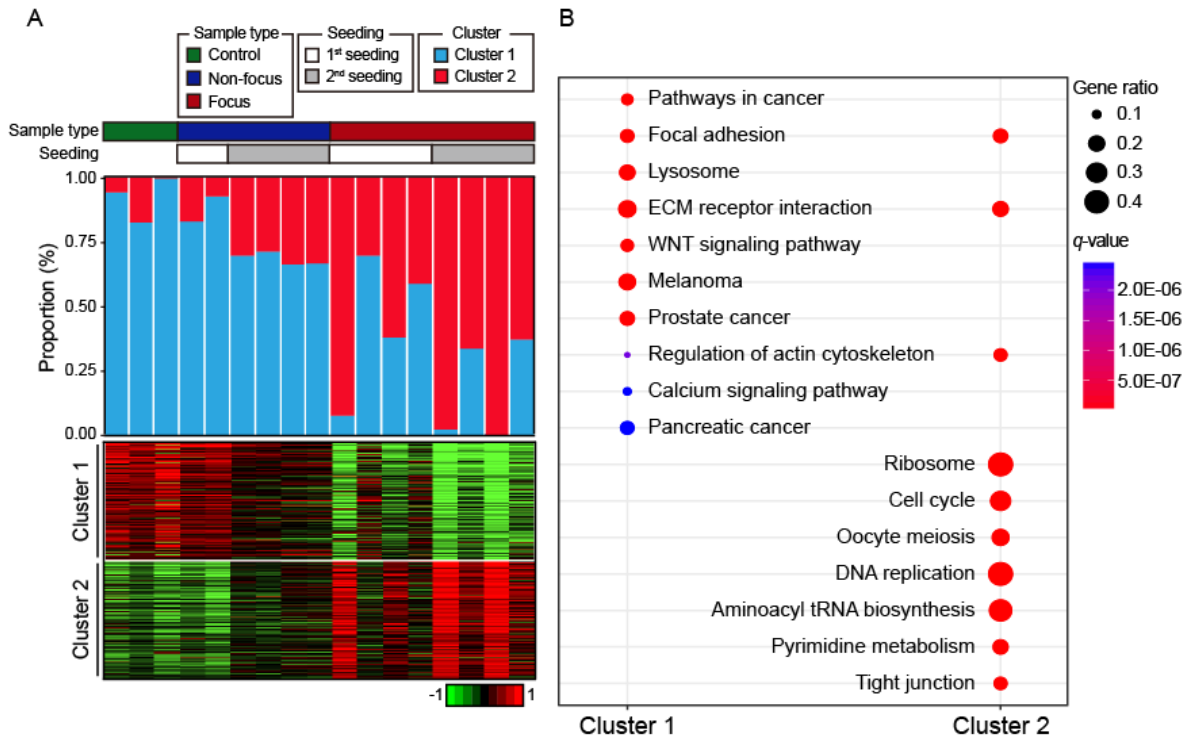

Figure 9. Estimation of cellular status of all samples. (A) The transcriptome profile of all samples. Two distinctive clusters represented the difference between gene expression pattern of samples with focus and without focus. (B) Dot plot illustrated Top 10 the most significant KEGG pathways that were enriched with each cluster.

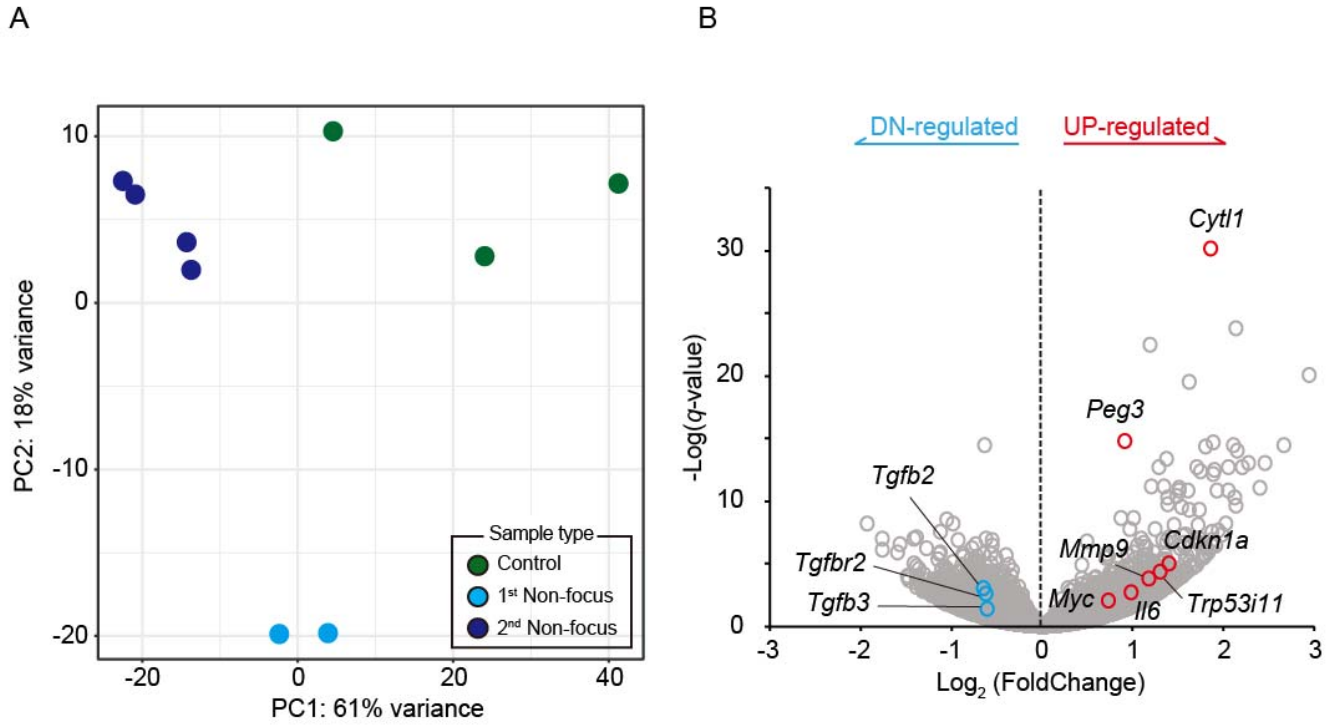

Figure 10. The transcriptome profile of Control and Non-focus. (A) Principal component analysis. Control samples were evidently separated from Non-focus samples. (B) Volcano plot described differentially expressed genes between Control and Non-focus samples.

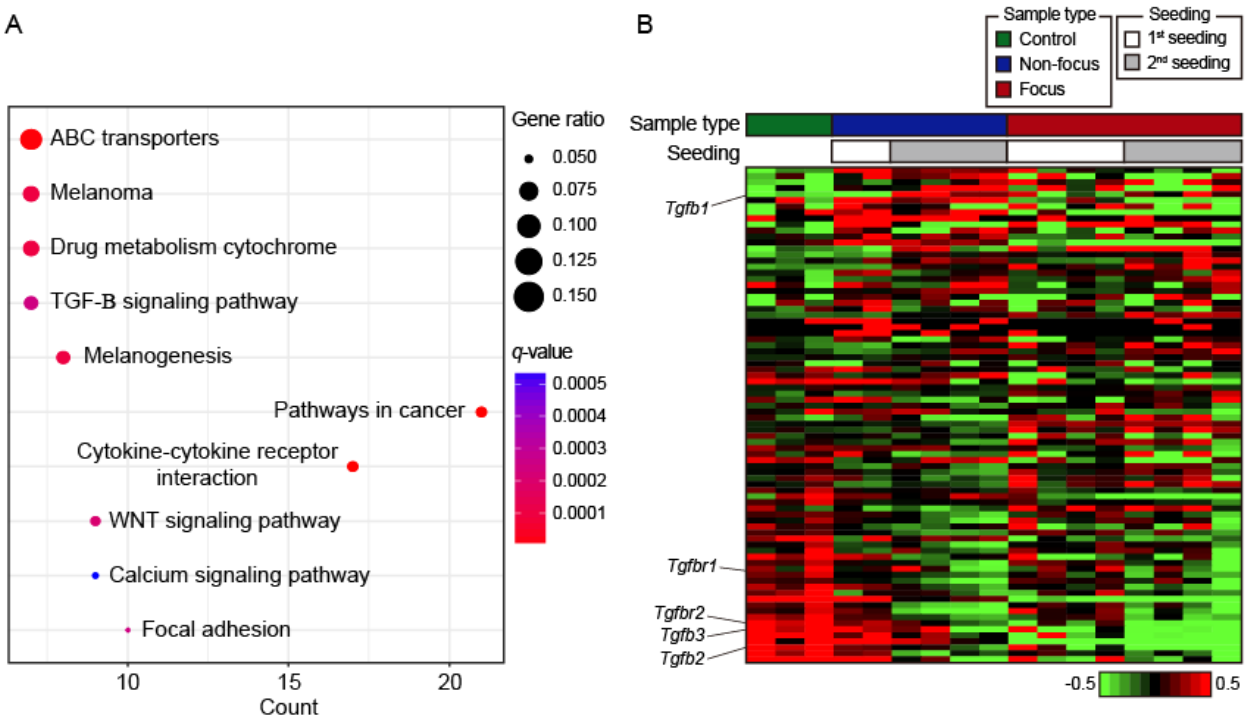

Figure 11. Down-regulated differentially expressed genes in Non-focus. (A) The top 10 the most significantly enriched KEGG pathways. (B) The heatmap exhibited gene expression pattern of TGF-β signaling pathway genes.

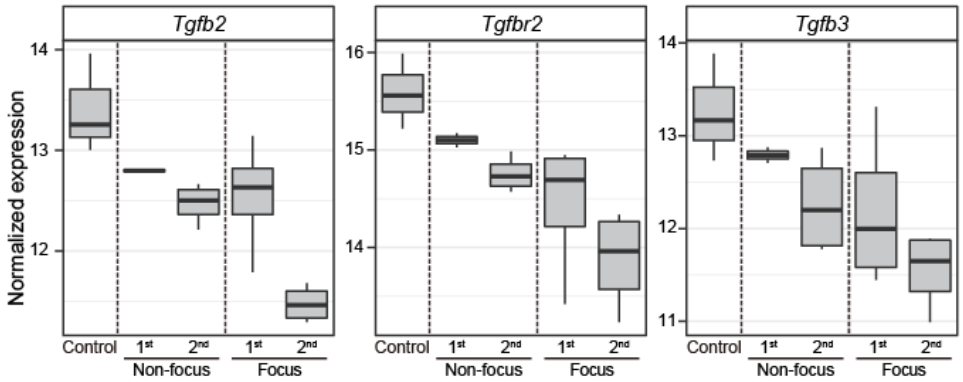

Figure 12. Relative expression levels of TGF-β genes. TGF-β genes showed down-regulation in Non-focus then further decrement was displayed in Focus.

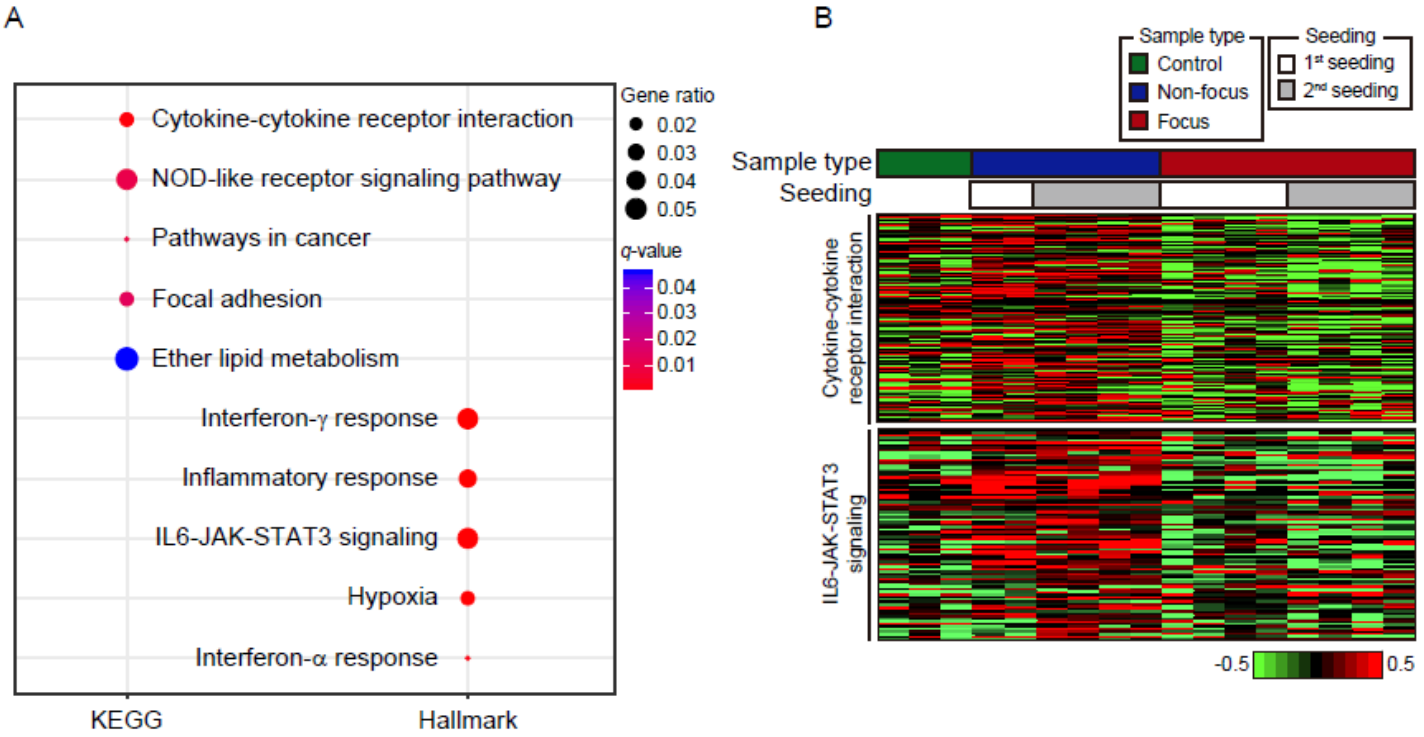

Figure 13. The characteristics of up-regulated genes in Non-focus samples when compared with Control. (A) Dot plot described the top 5 KEGG and HM pathways that were enriched with up-regulated genes. (B) The heatmap illustrated immune response related pathways were elevated in Non-focus samples.

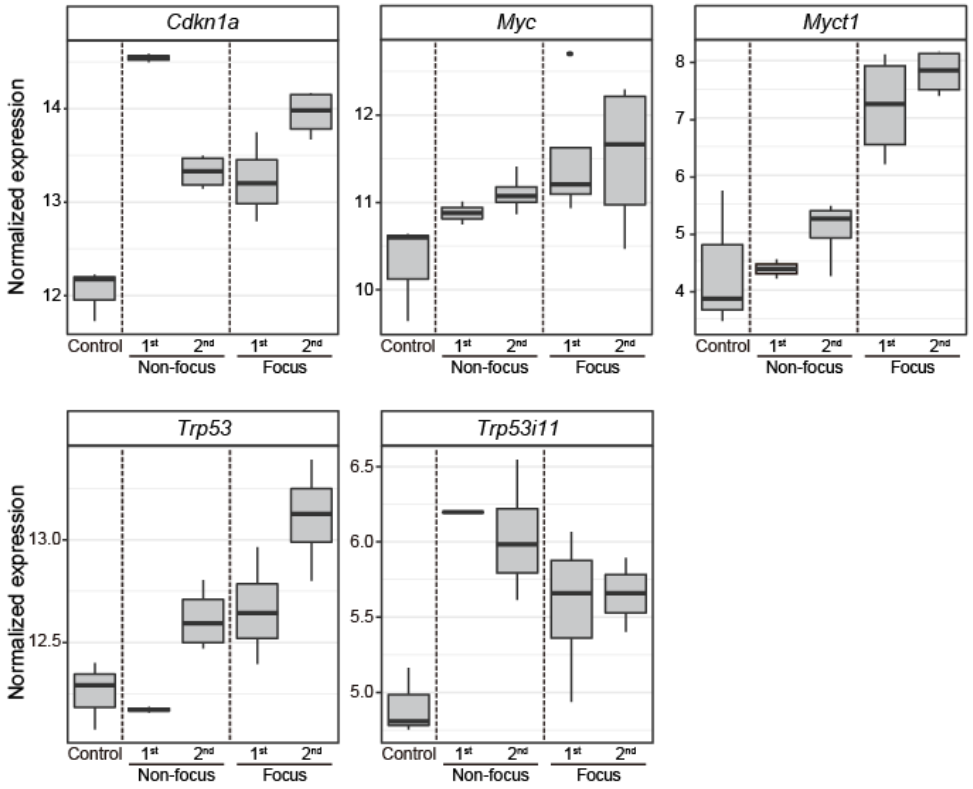

Figure 14. Relative expression levels of increased genes in Non-focus. Expression level of P53 pathway related genes were observed. Four genes except for *Cdkn1a* displayed gradual increment in Non-focus, then Focus. *Cdkn1a* exhibited dramatic up-regulation in 1st Non-focus, then showed slight decrement in 2nd Non-focus and Focus samples.

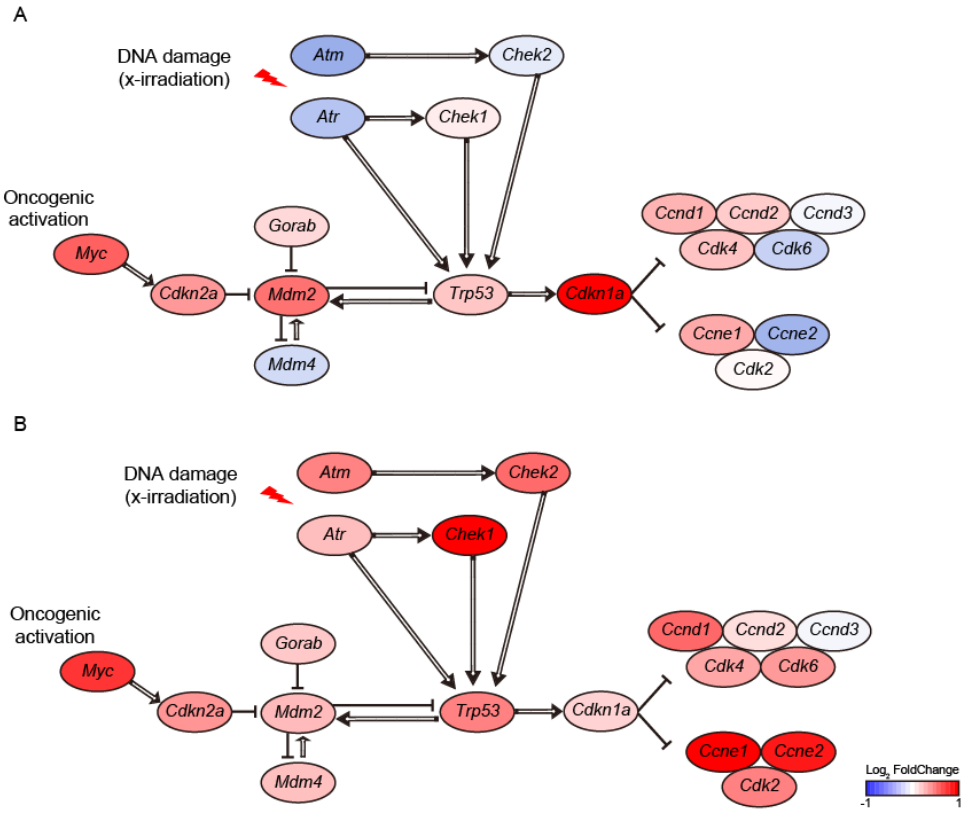

Figure 15. The schematic description of P53 signaling pathway. (A) Gene expression changes occurred in Non-focus samples when compared with Control and (B) in Focus samples when compared with Non-focus.

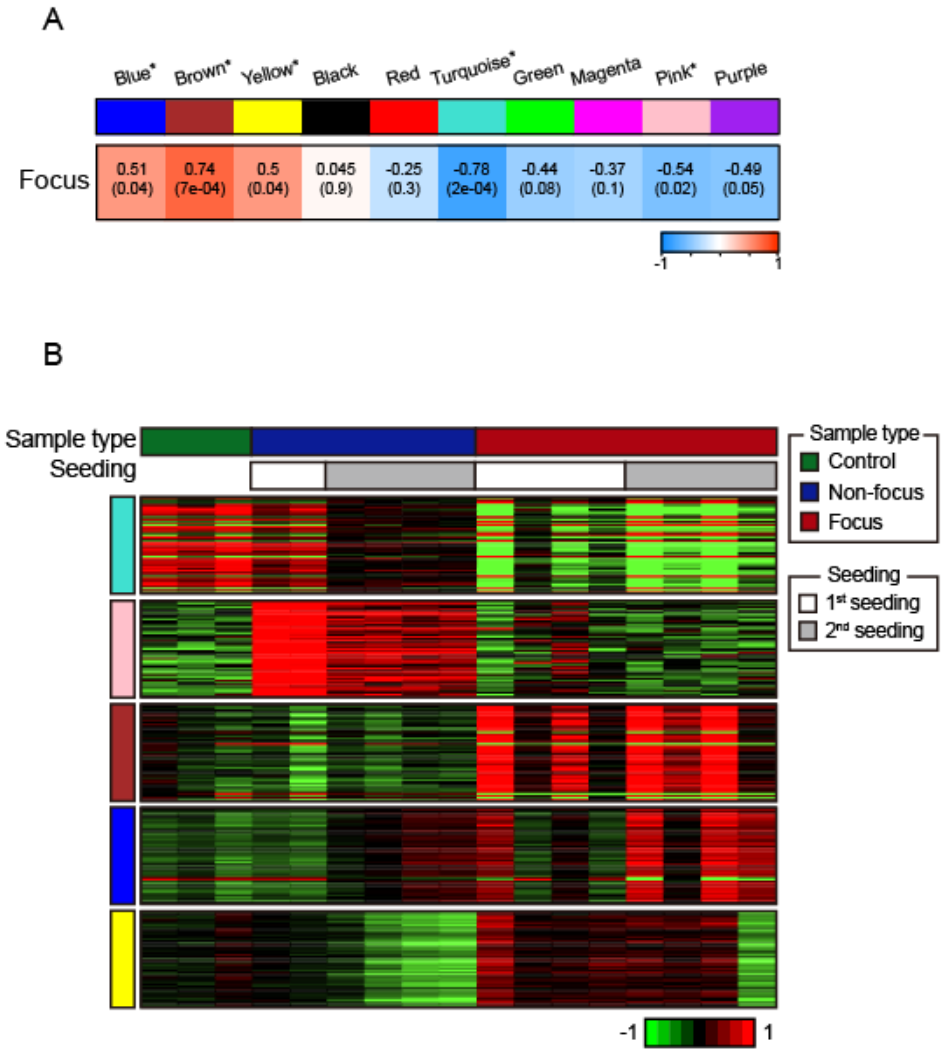

Figure 16. Weighted gene co-expression network analysis using all samples. (A) Module trait relationship of generated modules. The statistically significant modules were marked with asterisk. The correlation values and *p*-values (in the bracket) were shown. (B) The heatmap illustrated gene expression profile of five statistically significant modules in all samples.

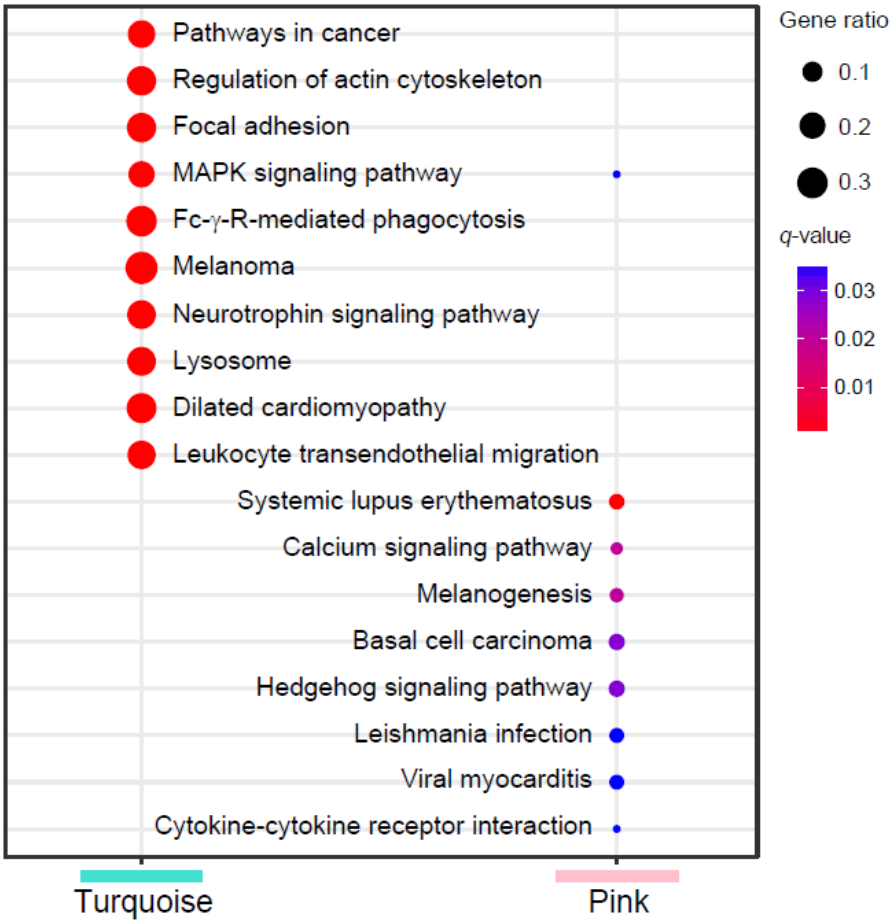

Figure 17. Pathway enrichment result for two significantly down-regulated modules in Focus. Top 10 the most enriched KEGG pathways were displayed.

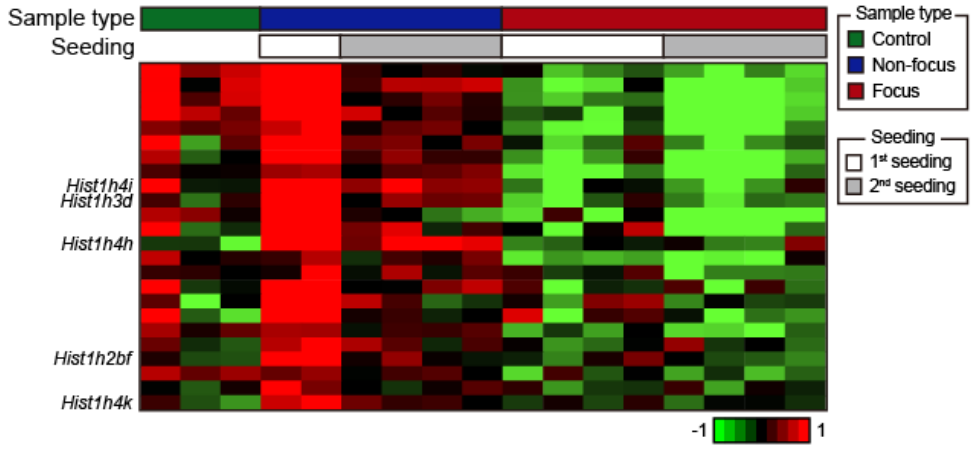

Figure 18. The expression pattern of histone genes in all samples. Focus samples showed decreased expression level of histone genes. The expression pattern of histone genes was observed. Marked five histone genes were found in pink module.

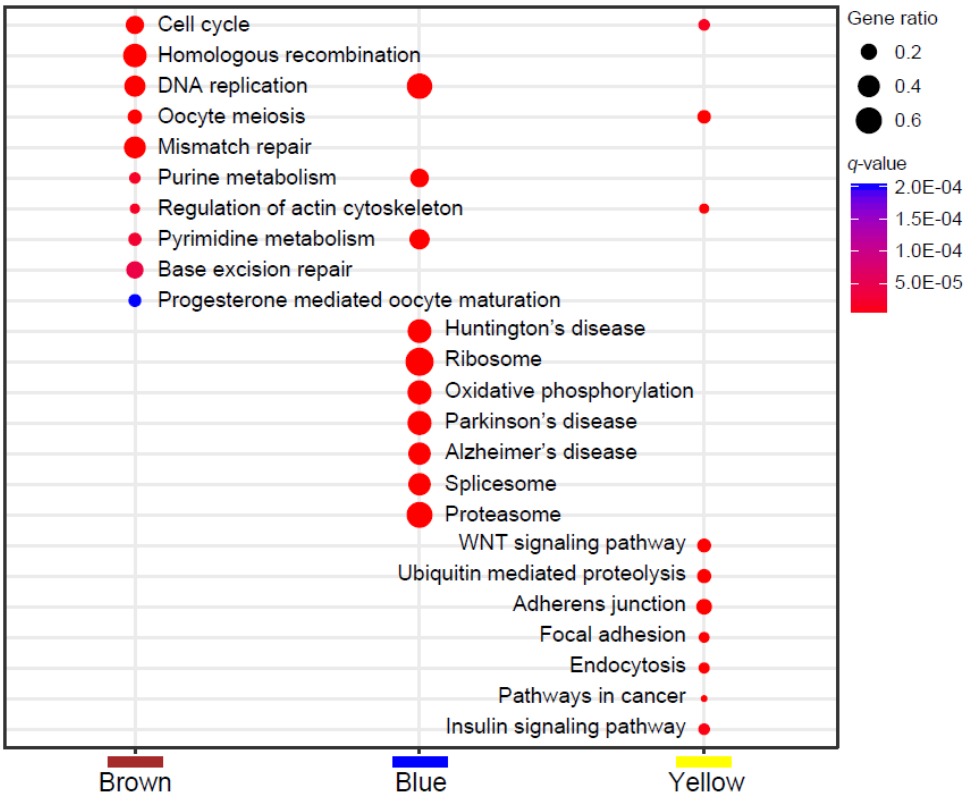

Figure 19. Pathway enrichment result for three significantly up-regulated modules in Focus. Top 10 the most enriched KEGG pathways were displayed.

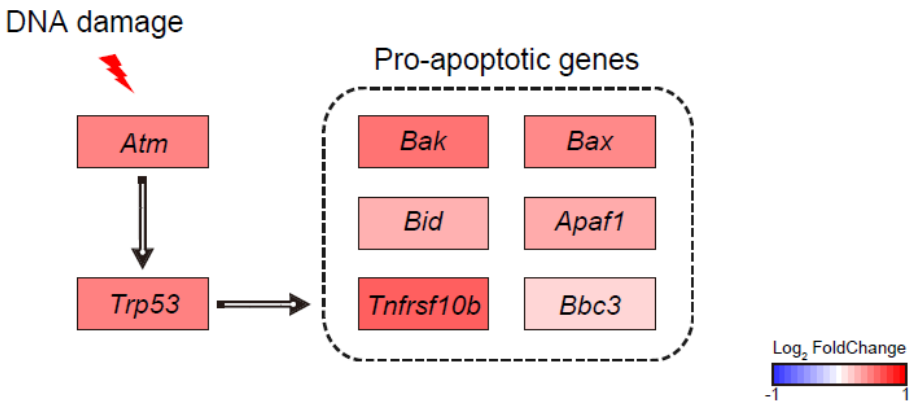

Figure 20. The schematic description of apoptotic pathway. Pro-apoptotic genes were up-regulated in Focus.

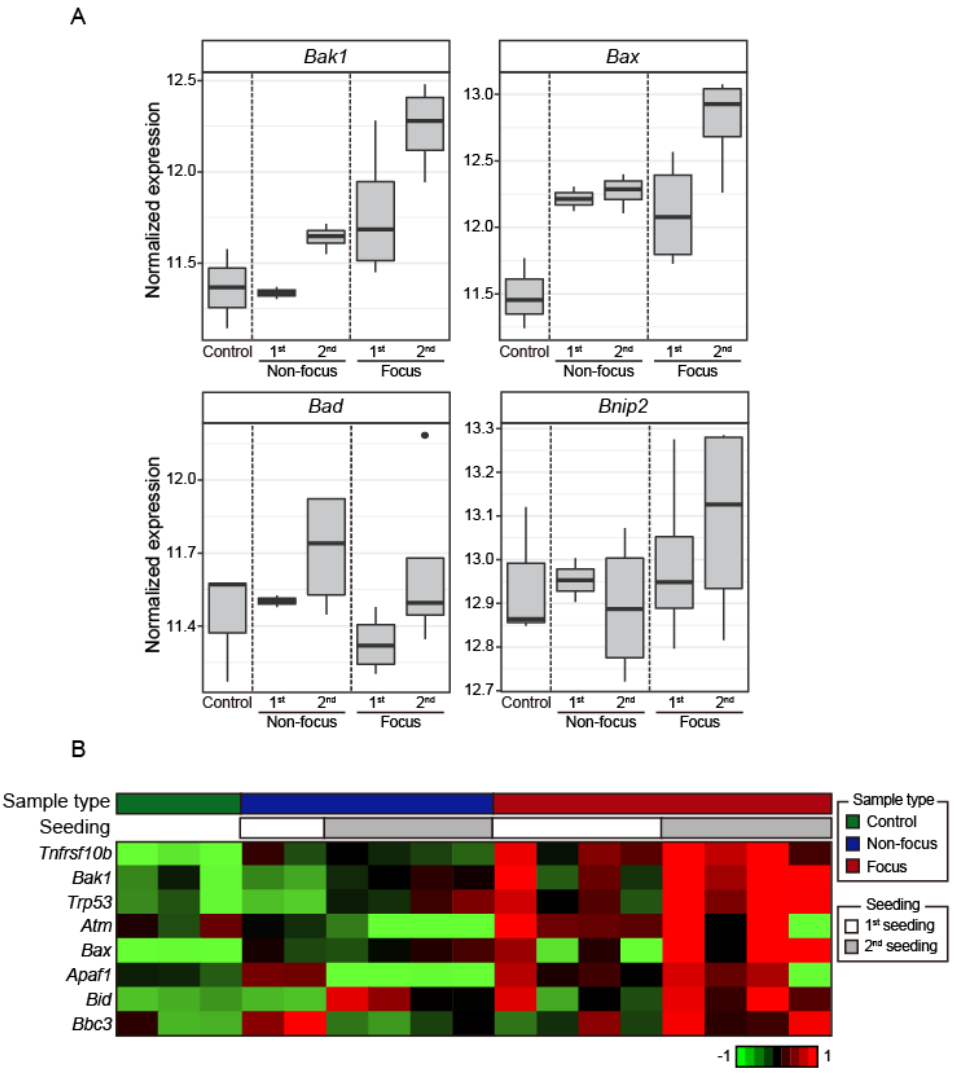

Figure 21. Transcriptome profile of apoptosis related genes. (A) Relative gene expression level of apoptosis related genes. (B) Gene expression pattern of apoptosis related genes in all samples.

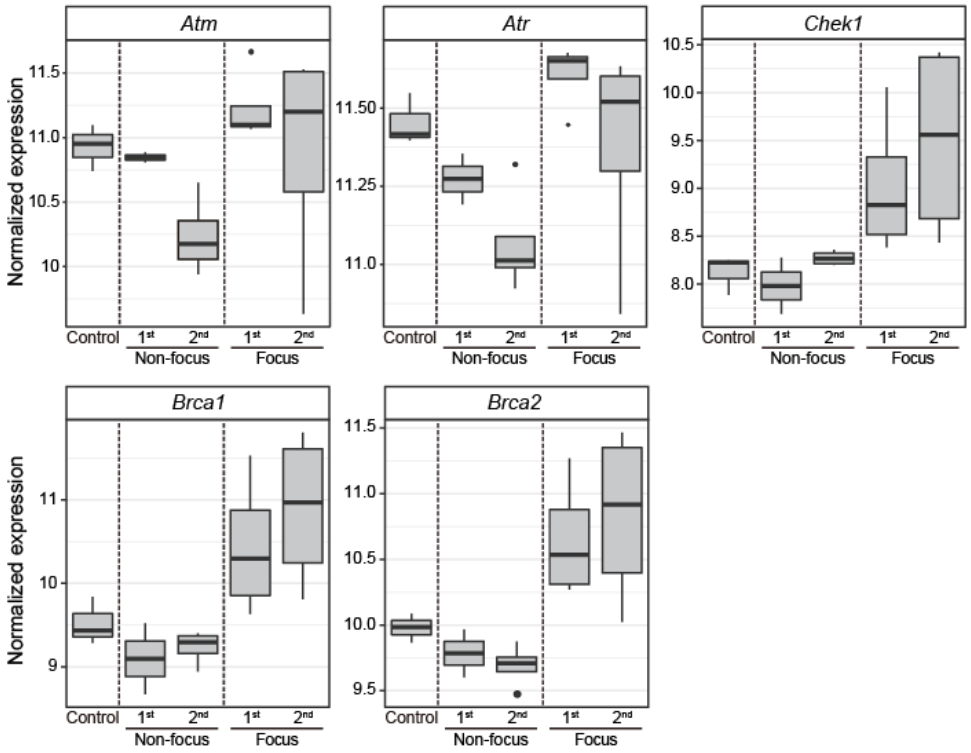

Figure 22. Relative expression level of DNA repair pathway related genes. DNA repair process related genes displayed up-regulation in Focus.

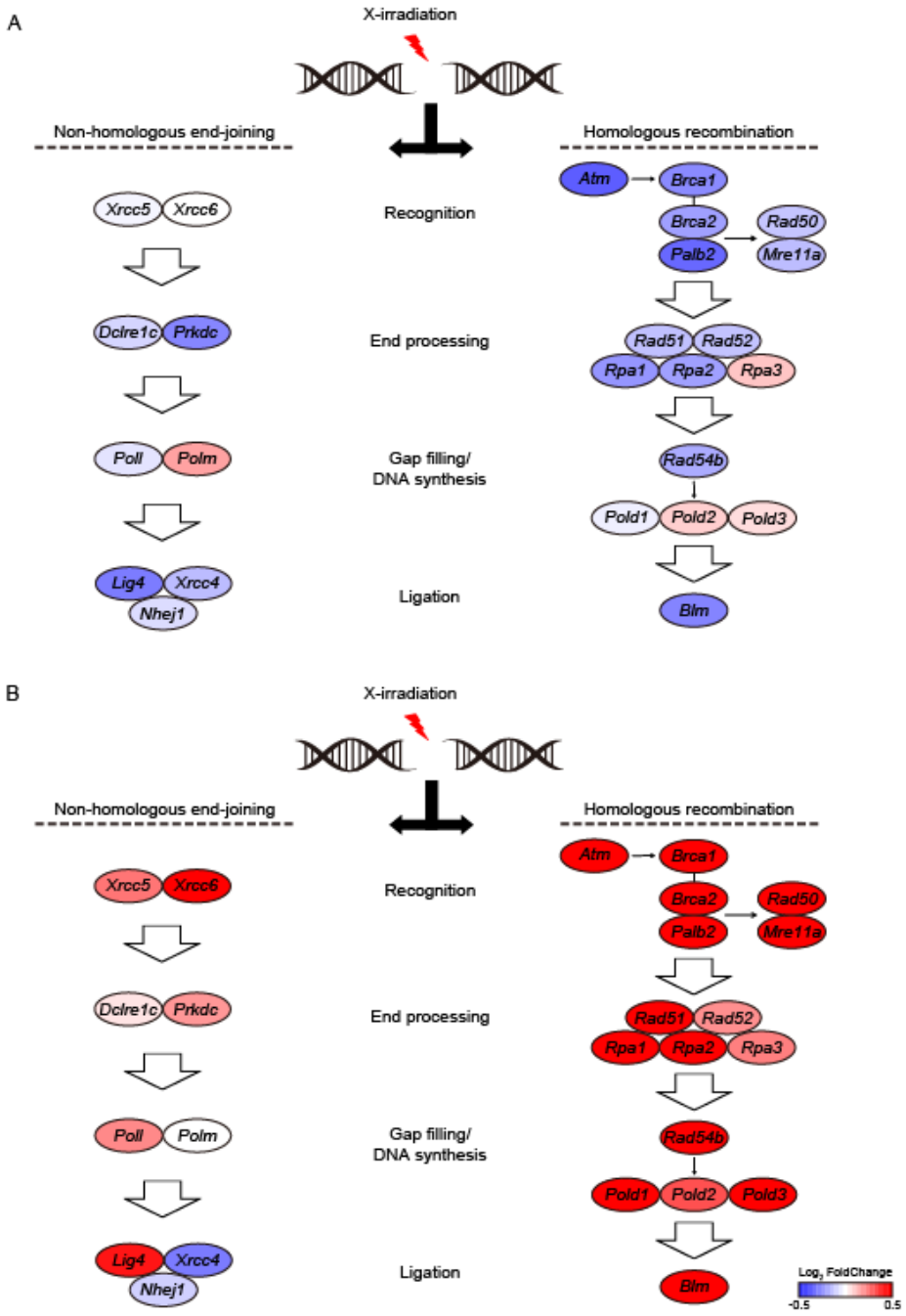

Figure 23. The schematic description of DNA repair pathways in (A) Non-focus and (B) Focus. Repair pathways were greatly up-regulated in Focus than Non-focus.

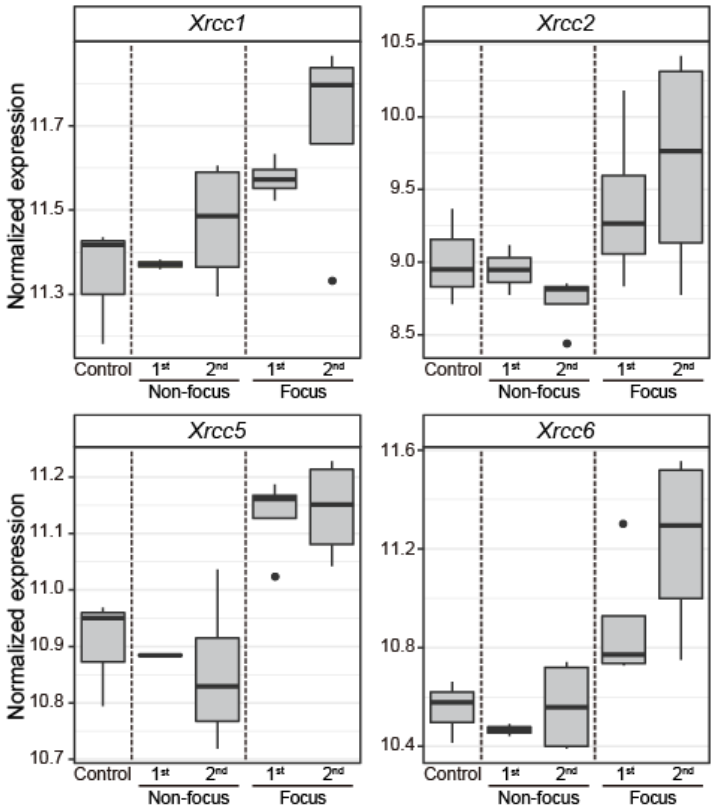

Figure 24. Relative expression level of *Xrcc* genes. Several *Xrcc* genes were elevated in Non-focus, then Focus and they were enriched with significantly up-regulated in Focus.

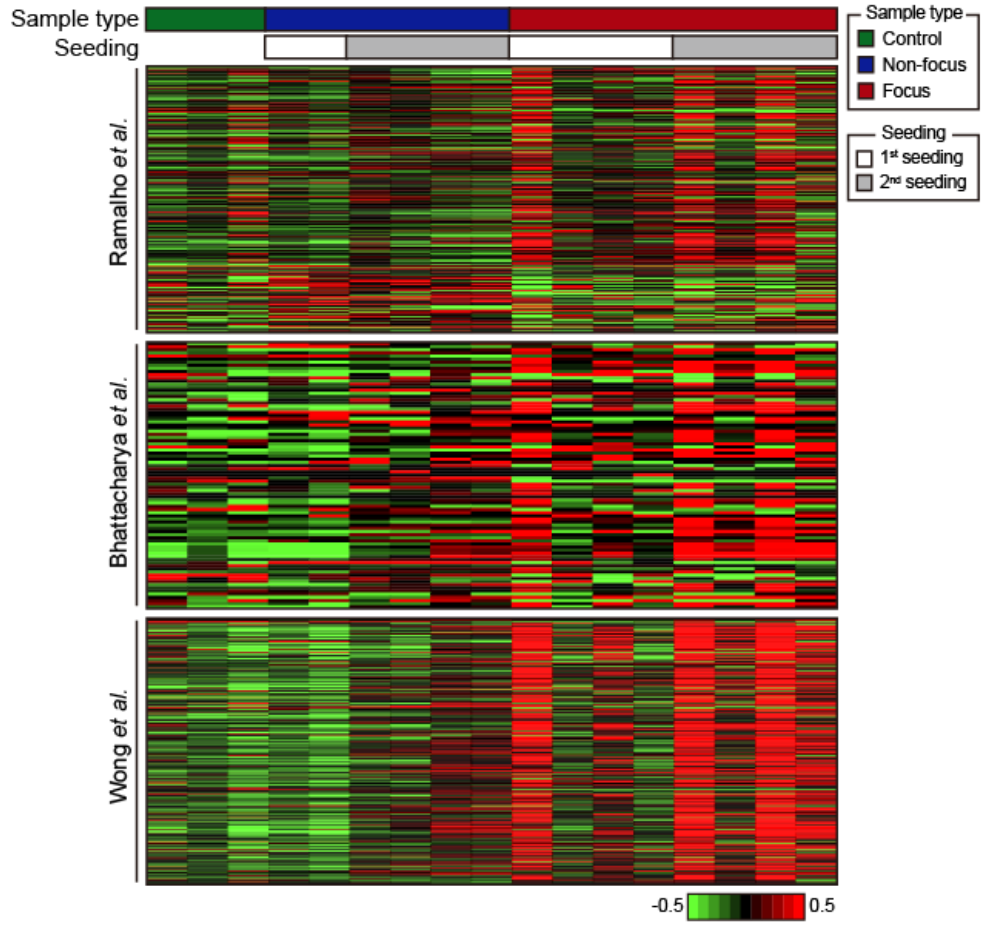

Figure 25. Expression profile of genes that describes *stemness* of transformed cells. Ramalho *et al.* represented 195 genes enriched in embryonic, neural, and hematopoietic stem cells and 69 genes depleted in embryonic, neural, and hematopoietic stem cells. Bhattacharya *et al.* discovered 85 genes showing stemness signature when up-regulated in six human embryonic stem cell (ESC) lines. Wong *et al.* found 328 core ESC-like genes that are up-regulated in mouse ESC which are shared with the human ESC-like genes. Focus cells exhibited stemness characteristics.

Discussion

Defining tumourigenesis always has been an important but challenging task. In mammalian cells, compromised genomic integrity might serve as an "ignition switch" of tumourigenesis. There are various known factors that initiate tumourigenesis although many of them are not fully elucidated yet. Based on previous reports, genomic instability is one of the traits that defines cancer [116, 117] and genomic instability refers to events such as SNPs, and chromosomal rearrangements. In cancer genome, genomic instability may occur prior to or as a consequence of malignant transformation. Here, using next-generation sequencing approach, I uncovered the underlying contributing factors of malignant transformation of x-irradiated cells.

Malignant transformation is a process whereby cells obtain the properties of tumour. There are several theories regarding this event, and one of them was suggested by John Little in 1980. He proposed two-step hypothesis, yet, was not able to describe both factors that were involved in the two-step transformation process. However, a new era of genomics and transcriptomics was broadly opened, the unknown factors that contributed to malignant transformation now may be able to arise on the surface. Vast amount of data provided by NGS technology were interpreted in this study and, as a first step, genomic alterations were sought including copy number variations and driver mutations.

Unexpectedly, genomic changes caused by x-irradiation were not severe. Although genomic instability is one of the traits that defines tumour, irradiation did

not directly initiate genomic instability [118]. However, irradiation generated defective cellular responses then induced DNA damage or gene expression alteration, which led to impaired cellular homeostasis. In this study, the cells damaged by x-irradiation, but due to intracellular heterogeneity, the damaging effects varied between cells. This phenomenon also is a great challenge in precision cancer therapy as human cancers also exhibit intratumoural heterogeneity [119]. Here, the mutational signature result signified the heterogeneity of x-irradiated cells. The sole demonstration of COSMIC mutational signature 5 from both Non-focus and Focus samples illustrated each cell possesses tumour-like characteristics but the level of DNA damage differed, hence resulted in mixed population of cells, displaying heterogeneity. Perhaps with single-cell level analysis, this problem may be solved and I may be able to obtain more detailed information.

Nevertheless, the formation of focus upon x-irradiation was very evident event and morphological changes were definite. *In vitro* result confirmed the existence of alterations in cells, although the major changes did not involve genomic alterations. Thus, I needed to look this matter with different angle and the answer was transcriptome analysis. Frequently, tumourigenesis resulted from collapsed homeostasis and it is also known as somatic evolution, it is crucial to fully understand the causes and the consequences of this process. In terms of cancer evolution, some studies reported that transcriptomic changes occur prior to genomic changes [61, 62]. Therefore, I focused on elucidating gene expression changes of Non-focus and Focus upon x-irradiation and showed several evidences which supported transcriptome changes may promoted malignant transformation upon x-irradiation.

Based on Little's hypothesis, due to the first factor, x-irradiation, malignant transformation is initiated and Non-focus stage was reached prior to generate focus. Then, by the second factor, although unknown, focus was generated. Here, I can speculate that Non-focus stage is equivalent to very early stage tumour cell as it presented potential to develop into focus which is a form of tumour. Unleashing the traits of Non-focus would finally give an answer to two-step malignant transformation process, hence I focused on deciphering the transcriptome characteristics of Non-focus.

One of the most interesting alterations in Non-focus when compared to Control was decreased expression of *Tgfβ* genes. *Tgfβ* genes have unusual aspect as they have dual roles depend on the tumour stage [120, 121]. It is also known as "*Tgfβ* paradox" since in early stage cancer, those genes play tumour suppressive role and induce growth arrest, but in later stage or advanced cancer, they act otherwise and promote tumour progression [120]. Down-regulation of *Tgfβ* genes was frequently observed in early stage cancer and their tumour suppressive role in early stage cancer is well-studied using mouse model system [80, 120, 122]. Various transgenes were used to generate mouse model of *Tgfβ* including dominant negative type 2 *Tgfβ* receptor (*Tgfβr2*^{DN}) and type 2 *Tgfβ* knockout receptor (*Tgfβr2*^{null}). In these mouse models, an increased occurrence of tumour emergence was observed which supported tumour suppressive role of *Tgfβ* genes. The results from mouse models were in parallel with gene expression profile of Non-focus which supported our previous hypothesis. Non-focus samples seemed to be at the very early stage of

malignant transformation, or tumourigenesis, and down-regulated *Tgfβ* genes can be the driving factor for progression of malignant transformation.

Moreover, another striking discovery was dramatic up-regulation of *Cdkn1a* in Non-focus, especially derived from 1st seeding plate. *Cdkn1a* is known as a DNA damage sensing gene and gets recruited to damaged sites and accumulated [91, 92]. For this reason, *Cdkn1a* sometimes used as DNA damage site detection marker [92]. Spatio-temporal dynamics of *Cdkn1a* is worth observing as the expression level of this gene in Non-focus cells from initially seeded plate was far more elevated than that of re-seeded ones. Several previously reported studies mentioned that the underlying mechanism for transformation of non-irradiated cells into malignant cells may be radiation-induced bystander effects [118]. One of the suggested evidence was p53-mediated bystander effect and upon α-particle exposure on human skin fibroblasts, three- to four-fold increment was observed in p53 protein and its downstream gene *Cdkn1a* in bystander cells [123, 124]. Based on this result, I was able to deduct that elevated *Cdkn1a* gene expression is one of the signatures of Non-focus, in the reported study, bystander cells.

As x-irradiation induced malignant transformation process, Focus samples reflected tumour cell characteristics. Immune system in response to irradiation damage, it can participate both in protection from tumourigenesis through immune surveillance or cause an inflammatory response hence contribute to tumour promoting environment. In Non-focus, immune response related genes were highly elevated to fight back from external attack. However, in Focus, down-regulation of

immune response was observed which can be interpreted as failed immune surveillance system. Here, once again, Focus samples exhibited tumour-like characteristics by showing collapsed immune surveillance system.

Furthermore, DNA repair pathways were highly elevated in Focus. Usually, in human cancer, DNA repair pathways show frequent down-regulation [125, 126]. However, in this study, in Focus, repair pathways, both HR and NHEJ, were highly up-regulated, especially HR. One of previous studies described the relationship between oncogenic transformation *in vitro* by x-irradiation and its effect on repair process [127]. At that time, the clear association between two events were not fully elucidated. The study only described that upon x-irradiation, the survival of cells was declined which implied dysfunction of repair process. In this study, I tried to clarify the influence of dysregulated DNA repair process and progression of malignant transformation. Although elevated DNA repair process is not frequent event, it is reported that, in *E. coli*, when error-free repair system is saturated, error-prone repair is triggered and it is called as the SOS repair system [17]. In addition, in *E. coli*, it is known that the SOS system is mainly responsible for UV radiation induced damage [17]. Moreover, also in human lymphoblast cells, upon radiation-induced DSBs generation, they are usually repaired by NHEJ, not by HR [128]. Even though the damage is repaired by HR, the yield is very low [129, 130]. This study displayed parallel results from *E. coli* and Little's experiment also showed the alternative error-prone DNA repair system was activated when error-free repair pathway was overloaded. As it displayed indirect tumour induction via x-irradiation, at that time, the involved process might have been quite obscure and vague. However, one thing

that was clear was the SOS system was permanently activated by the x-irradiation. Therefore, low yield error-free repair pathway and error-prone repair system were not able to fix damaged DNA properly, hence the Focus cells with mutagenic DNA lesions remained unrepaired and underwent uncontrolled cell division then showed tumour-like characteristics. The similar trend was exhibited in apoptotic pathway. The damaged cells usually undergo cell cycle arrest, then experience apoptosis. Here, through compensatory mechanism, Focus tried to undergo apoptosis to eliminate damaged cells. Since both DNA repair mechanism and apoptotic pathway were not working properly, no matter how hard the Focus cells tried to fight back for the damage, the mission did not succeed and as a consequence, focus were rigidly formed.

Based on findings from this study, transcriptome alterations contributed the most to two-step malignant transformation. However, gaining tumour-like malignancy characteristics during transformation process may need some additional explanations. Various studies reported that stem cell is involved in tumourigenesis. Stem cells are significant to tissue repair but they also have functions in malignant transformation and tumour initiation [109-112]. When ESC-like genes from CGP were analysed, Focus displayed similar expression pattern as ESCs which reinforced that Focus exhibits "stemness". Moreover, it is reported that increased *Myc* oncogene is also involved in reprogramming cells. In fact, *Myc* alone is sufficient to activate ESC-like programme in normal or tumour cells [115]. It is also reported that tumour initiated by *Myc* is independent of genomic instability. In this study, *Myc* is up-regulated in Focus and genomic alteration in Focus is very mild. Taken together,

"stemness" of Focus promoted the second step of two-step malignant transformation by reprogramming cells into cancer stem cell, hence allowed to gain malignancy in transformed cells. In conclusion, there are three factors involved in the second step of two-step malignant transformation process: down-regulated *Tgfβ* gene expression, activated error-prone DNA repair system, and "stemness" of Focus cells.

From genomic and transcriptomic analyses, I was able to conclude that x-irradiation induced focus generation represented tumourigenesis process *in vitro*, and the generated foci can be considered as very early stage cancer cells. I was able to identify the transcriptome profile of Non-focus and Focus and discovered the factors that contributed to the second step of Little's two-step hypothesis. Having full knowledge on malignant transformation process is important as knowing this process may be the initial step for developing more efficient cancer treatments. The targets for conventional cancer therapies were usually found with genomic studies. However, with our approach, as sometimes transcriptome changes occur prior to genomic alterations, I believe that cancer can be detected at its earlier stage with transcriptomic markers which will lead to better cure for cancer.

References

1. Loeb, L.A., K.R. Loeb, and J.P. Anderson, *Multiple mutations and cancer.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 776-81.
2. Tomlinson, I.P., M.R. Novelli, and W.F. Bodmer, *The mutation rate and cancer.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14800-3.
3. Weinstein, I.B. and K. Case, *The history of Cancer Research: introducing an AACR Centennial series.* Cancer Res, 2008. **68**(17): p. 6861-2.
4. Yamagiwa, K. and K. Ichikawa, *Experimental study of the pathogenesis of carcinoma.* CA Cancer J Clin, 1977. **27**(3): p. 174-81.
5. Ellerman, V. and O. Bang, *Experimentelle Leukämie bei Hühnern.* Zentralbl. Bakteriol. Parasitenkd. Infectionskr. Hyg. Abt. Orig. , 1908. **46**: p. 595-609.
6. Rous, P., *A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells.* J Exp Med, 1911. **13**(4): p. 397-411.
7. Rous, P., *Surmise and fact on the nature of cancer.* Nature, 1959. **183**(4672): p. 1357-61.
8. Kennedy, A.R., et al., *Relationship between x-ray exposure and malignant transformation in C3H 10T1/2 cells.* Proc Natl Acad Sci U S A, 1980. **77**(12): p. 7262-6.
9. Terzaghi, M. and J.B. Little, *X-radiation-induced transformation in a C3H mouse embryo-derived cell line.* Cancer Res, 1976. **36**(4): p. 1367-74.
10. Kennedy, A.R., J. Cairns, and J.B. Little, *Timing of the steps in transformation of C3H 10T 1/2 cells by X-irradiation.* Nature, 1984.

307(5946): p. 85-6.

11. Park, W.-Y., et al., *Effects of Phorbol Ester and Amphotericin B on the X-Radiation Induced Transformation of C3H 10T1/2 Cells.* Experimental and Molecular Medicine, 1992. **25**(2): p. 157-162.
12. Torgovnick, A. and B. Schumacher, *DNA repair mechanisms in cancer development and therapy.* Front Genet, 2015. **6**: p. 157.
13. Little, J.W. and D.W. Mount, *The SOS regulatory system of Escherichia coli.* Cell, 1982. **29**(1): p. 11-22.
14. Olsson, M. and T. Lindahl, *Repair of alkylated DNA in Escherichia coli. Methyl group transfer from O6-methylguanine to a protein cysteine residue.* J Biol Chem, 1980. **255**(22): p. 10569-71.
15. Janion, C., *Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli.* Int J Biol Sci, 2008. **4**(6): p. 338-44.
16. Krishna, S., S. Maslov, and K. Sneppen, *UV-induced mutagenesis in Escherichia coli SOS response: a quantitative model.* PLoS Comput Biol, 2007. **3**(3): p. e41.
17. Darnell, J.E., H.F. Lodish, and D. Baltimore, *Molecular cell biology*. 1986, New York: Scientific American Books : Distributed by W.H. Freeman. xxxv, 1187 p.
18. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology.* Trends Genet, 2014. **30**(9): p. 418-26.
19. Park, S.T. and J. Kim, *Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing.* Int Neurourol J, 2016. **20**(Suppl 2): p.

S76-83.

20. Behjati, S. and P.S. Tarpey, *What is next generation sequencing?* Arch Dis Child Educ Pract Ed, 2013. **98**(6): p. 236-8.

21. Seo, J.S., et al., *De novo assembly and phasing of a Korean human genome.* Nature, 2016. **538**(7624): p. 243-247.

22. Trapnell, C. and S.L. Salzberg, *How to map billions of short reads onto genomes.* Nat Biotechnol, 2009. **27**(5): p. 455-7.

23. Ekblom, R. and J.B. Wolf, *A field guide to whole-genome sequencing, assembly and annotation.* Evol Appl, 2014. **7**(9): p. 1026-42.

24. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

25. Roberts, A., et al., *Identification of novel transcripts in annotated genomes using RNA-Seq.* Bioinformatics, 2011. **27**(17): p. 2325-9.

26. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

27. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

28. Li, Y. and T.O. Tollefsbol, *DNA methylation detection: bisulfite genomic sequencing analysis.* Methods Mol Biol, 2011. **791**: p. 11-21.

29. Vidal, E., et al., *A DNA methylation map of human cancer at single base-pair resolution.* Oncogene, 2017. **36**(40): p. 5648-5657.

30. Lim, B.C., et al., *Hoyeraal-Hreidarsson syndrome with a DKC1 mutation identified by whole-exome sequencing.* Gene, 2014. **546**(2): p. 425-9.

31. Yoo, S.K., et al., *Noninvasive prenatal diagnosis of duchenne muscular dystrophy: comprehensive genetic diagnosis in carrier, proband, and fetus.* Clin Chem, 2015. **61**(6): p. 829-37.

32. Little, J.B., *Radiation-induced genomic instability.* Int J Radiat Biol, 1998. **74**(6): p. 663-71.

33. Huang, L., A.R. Snyder, and W.F. Morgan, *Radiation-induced genomic instability and its implications for radiation carcinogenesis.* Oncogene, 2003. **22**(37): p. 5848-54.

34. Barcellos-Hoff, M.H. and A.L. Brooks, *Extracellular signaling through the microenvironment: a hypothesis relating carcinogenesis, bystander effects, and genomic instability.* Radiat Res, 2001. **156**(5 Pt 2): p. 618-27.

35. Baverstock, K., *Radiation-induced genomic instability: a paradigm-breaking phenomenon and its relevance to environmentally induced cancer.* Mutat Res, 2000. **454**(1-2): p. 89-109.

36. Puck, T.T., P.I. Marcus, and S.J. Cieciura, *Clonal growth of mammalian cells in vitro; growth characteristics of colonies from single HeLa cells with and without a feeder layer.* J Exp Med, 1956. **103**(2): p. 273-83.

37. Engstrom, P.G., et al., *Systematic evaluation of spliced alignment programs for RNA-seq data.* Nat Methods, 2013. **10**(12): p. 1185-91.

38. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

39. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

40. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **43**: p. 11 10 1-33.

41. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

42. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.

43. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

44. Talevich, E., et al., *CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing.* PLoS Comput Biol, 2016. **12**(4): p. e1004873.

45. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

46. Dey, K.K., C.J. Hsiao, and M. Stephens, *Visualizing the structure of RNA-seq expression data using grade of membership models.* PLoS Genet, 2017. **13**(3): p. e1006599.

47. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population*

structure using multilocus genotype data. Genetics, 2000. **155**(2): p. 945-59.

48. Lin, S.M., et al., *Model-based variance-stabilizing transformation for Illumina microarray data.* Nucleic Acids Res, 2008. **36**(2): p. e11.

49. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.

50. Cho, B.A., et al., *Transcriptome Network Analysis Reveals Aging-Related Mitochondrial and Proteasomal Dysfunction and Immune Activation in Human Thyroid.* Thyroid, 2018. **28**(5): p. 656-666.

51. Kanehisa, M., et al., *KEGG as a reference resource for gene and protein annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D457-62.

52. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.

53. Olshen, A.B., et al., *Parent-specific copy number in paired tumor-normal studies using circular binary segmentation.* Bioinformatics, 2011. **27**(15): p. 2038-46.

54. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the analysis of array CGH data.* Bioinformatics, 2007. **23**(6): p. 657-63.

55. Rausch, T., et al., *DELLY: structural variant discovery by integrated paired-end and split-read analysis.* Bioinformatics, 2012. **28**(18): p. i333-i339.

56. Gehring, J.S., et al., *SomaticSignatures: inferring mutational signatures from single-nucleotide variants.* Bioinformatics, 2015. **31**(22): p. 3673-5.

57. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.*

Nature, 2013. **500**(7463): p. 415-21.

58. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.

59. Willis, N.A., E. Rass, and R. Scully, *Deciphering the Code of the Cancer Genome: Mechanisms of Chromosome Rearrangement.* Trends Cancer, 2015. **1**(4): p. 217-230.

60. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

61. Podlaha, O., et al., *Evolution of the cancer genome.* Trends Genet, 2012. **28**(4): p. 155-63.

62. Lowdon, R.F., H.S. Jang, and T. Wang, *Evolution of Epigenetic Regulation in Vertebrate Genomes.* Trends Genet, 2016. **32**(5): p. 269-283.

63. Nik-Zainal, S., et al., *Mutational processes molding the genomes of 21 breast cancers.* Cell, 2012. **149**(5): p. 979-93.

64. Alexandrov, L.B., et al., *Deciphering signatures of mutational processes operative in human cancer.* Cell Rep, 2013. **3**(1): p. 246-59.

65. Helleday, T., S. Eshtad, and S. Nik-Zainal, *Mechanisms underlying mutational signatures in human cancers.* Nat Rev Genet, 2014. **15**(9): p. 585-98.

66. Alexandrov, L.B. and M.R. Stratton, *Mutational signatures: the patterns of somatic mutations hidden in cancer genomes.* Curr Opin Genet Dev, 2014. **24**: p. 52-60.

67. Dang, C.V., *MYC on the path to cancer.* Cell, 2012. **149**(1): p. 22-35.

68. Lin, C.Y., et al., *Transcriptional amplification in tumor cells with elevated c-Myc.* Cell, 2012. **151**(1): p. 56-67.

69. Gabay, M., Y. Li, and D.W. Felsher, *MYC activation is a hallmark of cancer initiation and maintenance.* Cold Spring Harb Perspect Med, 2014. **4**(6).

70. Escot, C., et al., *Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas.* Proc Natl Acad Sci U S A, 1986. **83**(13): p. 4834-8.

71. Ladanyi, M., et al., *Sporadic amplification of the MYC gene in human osteosarcomas.* Diagn Mol Pathol, 1993. **2**(3): p. 163-7.

72. Gamberi, G., et al., *C-myc and c-fos in human osteosarcoma: prognostic value of mRNA and protein expression.* Oncology, 1998. **55**(6): p. 556-63.

73. Kawate, S., et al., *Amplification of c-myc in hepatocellular carcinoma: correlation with clinicopathologic features, proliferative activity and p53 overexpression.* Oncology, 1999. **57**(2): p. 157-63.

74. Stock, C., et al., *Chromosomal regions involved in the pathogenesis of osteosarcomas.* Genes Chromosomes Cancer, 2000. **28**(3): p. 329-36.

75. Boxer, L.M. and C.V. Dang, *Translocations involving c-myc and c-myc function.* Oncogene, 2001. **20**(40): p. 5595-610.

76. Gialeli, C., A.D. Theocharis, and N.K. Karamanos, *Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting.* FEBS J, 2011. **278**(1): p. 16-27.

77. Tatti, O., et al., *MMP16 Mediates a Proteolytic Switch to Promote Cell-Cell Adhesion, Collagen Alignment, and Lymphatic Invasion in Melanoma.*

Cancer Res, 2015. **75**(10): p. 2083-94.

78. Sun, Q., et al., *MMP25 (MT6-MMP) is highly expressed in human colon cancer, promotes tumor growth, and exhibits unique biochemical properties.* J Biol Chem, 2007. **282**(30): p. 21998-2010.

79. Jian, P., et al., *MMP28 (epilysin) as a novel promoter of invasion and metastasis in gastric cancer.* BMC Cancer, 2011. **11**: p. 200.

80. Derynck, R., R.J. Akhurst, and A. Balmain, *TGF-beta signaling in tumor suppression and cancer progression.* Nat Genet, 2001. **29**(2): p. 117-29.

81. Tanaka, T., M. Narazaki, and T. Kishimoto, *IL-6 in inflammation, immunity, and disease.* Cold Spring Harb Perspect Biol, 2014. **6**(10): p. a016295.

82. Glasset, B., et al., *Bacillus cereus, a serious cause of nosocomial infections: Epidemiologic and genetic survey.* PLoS One, 2018. **13**(5): p. e0194346.

83. Relaix, F., et al., *Pw1/Peg3 is a potential cell death mediator and cooperates with Siah1a in p53-mediated apoptosis.* Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2105-10.

84. Shay, G., C.C. Lynch, and B. Fingleton, *Moving targets: Emerging roles for MMPs in cancer progression and metastasis.* Matrix Biol, 2015. **44-46**: p. 200-6.

85. Stine, Z.E., et al., *MYC, Metabolism, and Cancer.* Cancer Discov, 2015. **5**(10): p. 1024-39.

86. Cazzalini, O., et al., *Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response.* Mutat Res, 2010. **704**(1-3): p. 12-20.

87. Hanova, M., et al., *DNA damage, DNA repair rates and mRNA expression*

levels of cell cycle genes (TP53, p21(CDKN1A), BCL2 and BAX) with respect to occupational exposure to styrene. Carcinogenesis, 2011. **32**(1): p. 74-9.

88. Stivala, L.A. and E. Prosperi, *Analysis of p21CDKN1A recruitment to DNA excision repair foci in the UV-induced DNA damage response.* Methods Mol Biol, 2004. **281**: p. 73-89.

89. Swann, J.B. and M.J. Smyth, *Immune surveillance of tumors.* J Clin Invest, 2007. **117**(5): p. 1137-46.

90. Grivennikov, S.I., F.R. Greten, and M. Karin, *Immunity, inflammation, and cancer.* Cell, 2010. **140**(6): p. 883-99.

91. Perucca, P., et al., *Spatiotemporal dynamics of p21CDKN1A protein recruitment to DNA-damage sites and interaction with proliferating cell nuclear antigen.* J Cell Sci, 2006. **119**(Pt 8): p. 1517-27.

92. Koike, M., Y. Yutoku, and A. Koike, *Accumulation of p21 proteins at DNA damage sites independent of p53 and core NHEJ factors following irradiation.* Biochem Biophys Res Commun, 2011. **412**(1): p. 39-43.

93. Sherr, C.J. and J.M. Roberts, *CDK inhibitors: positive and negative regulators of G1-phase progression.* Genes Dev, 1999. **13**(12): p. 1501-12.

94. Dotto, G.P., *p21(WAF1/Cip1): more than a break to the cell cycle?* Biochim Biophys Acta, 2000. **1471**(1): p. M43-56.

95. Slattery, M.L., et al., *The p53-signaling pathway and colorectal cancer: Interactions between downstream p53 target genes and miRNAs.* Genomics, 2018.

96. Su, C., et al., *DNA damage induces downregulation of histone gene expression through the G1 checkpoint pathway.* EMBO J, 2004. **23**(5): p. 1133-43.

97. Marechal, A. and L. Zou, *DNA damage sensing by the ATM and ATR kinases.* Cold Spring Harb Perspect Biol, 2013. **5**(9).

98. Enriquez-Rios, V., et al., *DNA-PKcs, ATM, and ATR Interplay Maintains Genome Integrity during Neurogenesis.* J Neurosci, 2017. **37**(4): p. 893-905.

99. Thacker, J. and M.Z. Zdzienicka, *The mammalian XRCC genes: their roles in DNA repair and genetic stability.* DNA Repair (Amst), 2003. **2**(6): p. 655-72.

100. Thompson, L.H., et al., *Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange.* Mol Cell Biol, 1990. **10**(12): p. 6160-71.

101. Schwartz, J.L., S. Giovanazzi, and R.R. Weichselbaum, *Recovery from sublethal and potentially lethal damage in an X-ray-sensitive CHO cell.* Radiat Res, 1987. **111**(1): p. 58-67.

102. Jones, N.J., R. Cox, and J. Thacker, *Isolation and cross-sensitivity of X-ray-sensitive mutants of V79-4 hamster cells.* Mutat Res, 1987. **183**(3): p. 279-86.

103. Tambini, C.E., et al., *The XRCC2 DNA repair gene: identification of a positional candidate.* Genomics, 1997. **41**(1): p. 84-92.

104. Liu, N., et al., *XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and*

other damages. Mol Cell, 1998. **1**(6): p. 783-93.

105. Thacker, J., *A surfeit of RAD51-like genes?* Trends Genet, 1999. **15**(5): p. 166-8.

106. Deans, B., et al., *Xrcc2 is required for genetic stability, embryonic neurogenesis and viability in mice.* EMBO J, 2000. **19**(24): p. 6675-85.

107. Tuteja, R. and N. Tuteja, *Ku autoantigen: a multifunctional DNA-binding protein.* Crit Rev Biochem Mol Biol, 2000. **35**(1): p. 1-33.

108. Nussenzweig, A., et al., *Requirement for Ku80 in growth and immunoglobulin V(D)J recombination.* Nature, 1996. **382**(6591): p. 551-5.

109. Torsvik, A., et al., *Spontaneous malignant transformation of human mesenchymal stem cells reflects cross-contamination: putting the research field on track - letter.* Cancer Res, 2010. **70**(15): p. 6393-6.

110. Serakinci, N., et al., *Adult human mesenchymal stem cell as a target for neoplastic transformation.* Oncogene, 2004. **23**(29): p. 5095-8.

111. Houghton, J., et al., *Gastric cancer originating from bone marrow-derived cells.* Science, 2004. **306**(5701): p. 1568-71.

112. Mohseny, A.B., et al., *Osteosarcoma originates from mesenchymal stem cells in consequence of aneuploidization and genomic loss of Cdkn2.* J Pathol, 2009. **219**(3): p. 294-305.

113. Ramalho-Santos, M., et al., *"Stemness": transcriptional profiling of embryonic and adult stem cells.* Science, 2002. **298**(5593): p. 597-600.

114. Bhattacharya, B., et al., *Gene expression in human embryonic stem cell lines: unique molecular signature.* Blood, 2004. **103**(8): p. 2956-64.

115. Wong, D.J., et al., *Module map of stem cell genes guides creation of epithelial cancer stem cells.* Cell Stem Cell, 2008. **2**(4): p. 333-44.

116. Shen, Z., *Genomic instability and cancer: an introduction.* J Mol Cell Biol, 2011. **3**(1): p. 1-3.

117. Yao, Y. and W. Dai, *Genomic Instability and Cancer.* J Carcinog Mutagen, 2014. **5**.

118. Lorimore, S.A., P.J. Coates, and E.G. Wright, *Radiation-induced genomic instability and bystander effects: inter-related nontargeted effects of exposure to ionizing radiation.* Oncogene, 2003. **22**(45): p. 7058-69.

119. Sabaawy, H.E., *Genetic Heterogeneity and Clonal Evolution of Tumor Cells and their Impact on Precision Cancer Medicine.* J Leuk (Los Angel), 2013. **1**(4): p. 1000124.

120. Principe, D.R., et al., *TGF-beta: duality of function between tumor prevention and carcinogenesis.* J Natl Cancer Inst, 2014. **106**(2): p. djt369.

121. Lebrun, J.-J., *The Dual Role of TGF in Human Cancer: From Tumor Suppression to Cancer Metastasis.* ISRN Molecular Biology, 2012. **2012**: p. 28.

122. Kubiczkova, L., et al., *TGF-beta - an excellent servant but a bad master.* J Transl Med, 2012. **10**: p. 183.

123. Azzam, E.I., et al., *Intercellular communication is involved in the bystander regulation of gene expression in human cells exposed to very low fluences of alpha particles.* Radiat Res, 1998. **150**(5): p. 497-504.

124. Azzam, E.I., S.M. de Toledo, and J.B. Little, *Direct evidence for the*

participation of gap junction-mediated intercellular communication in the transmission of damage signals from alpha -particle irradiated to nonirradiated cells. Proc Natl Acad Sci U S A, 2001. **98**(2): p. 473-8.

125. van Vugt, M.A., *Shutting down the power supply for DNA repair in cancer cells.* J Cell Biol, 2017. **216**(2): p. 295-297.

126. Bindra, R.S., et al., *Down-regulation of Rad51 and decreased homologous recombination in hypoxic cancer cells.* Mol Cell Biol, 2004. **24**(19): p. 8504-18.

127. Terzaghi, M. and J.B. Litle, *Oncogenic transformation in vitro by x-rays: influence of repair processes*. 1976, United States: Raven Press.

128. Benjamin, M.B. and J.B. Little, *X rays induce interallelic homologous recombination at the human thymidine kinase gene.* Mol Cell Biol, 1992. **12**(6): p. 2730-8.

129. Little, J.B., *Radiation carcinogenesis.* Carcinogenesis, 2000. **21**(3): p. 397-404.

130. Singleton, B.K., et al., *The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit.* Mol Cell Biol, 1999. **19**(5): p. 3267-77.

국문초록

X-선 조사 C3H10T1/2 세포주에서 단계별 형질전환의 유전체 및 전사체 연구

서울대학교 의과대학 의과학과

조 벼 리

종양의 발생과 진화는 다양한 종류의 변이에 의해 활성화 된다. 1980년, Little과 Kennedy는 종양으로 되는 단계인 세포의 형질전환을 초래하는 요인들을 밝히고자 C3H10T1/2 세포주에서 엑스레이를 이용하여 실험을 진행하였다. Little은 two-step 이론을 제시하며 세포에서 x-선 조사로 인해 유도되는 세포의 형질전환은 두 종류의 인자에 의해 일어난다고 제안하였다. 두 종류의 인자 중 첫 번째 인자는 x-선 조사였으나 두 번째 인자에 대해서는 자세하게 밝혀내지 못하였다. 40년의 세월이 지난 오늘날, 본 연구에서는 차세대 염기서열 분석 기법을 이용하여 1980년대에 미처 밝히지 못했던 형질전환을 초래하는 유전자군에 대해 밝히고자 하였다. 먼저, 전장 유전체 분석 기법을 이용하여 x-선 조사로 유도된 focus 형성과 유전형질의 변환과의 관계를 알아보고자 하였다. X-선이 조사 되었으나 focus를 형성하지 않은 세포 (Non-focus) 와 focus

가 형성된 세포 (Focus) 의 유전체 복제수 변이, 돌연변이 발생 원인에 따른 고유한 특징 및 종양변이부담을 분석하였다. 그 결과 focus를 형성하더라도 x-선이 조사된 focus를 형성하지 않은 세포에 비해 유전체의 변화는 크게 없는 것이 관찰되었으며 x-선 조사가 유전체 변이에 큰 영향은 주지 않는 것을 알 수 있었다. 그리하여 본 연구에서는 전사체 분석 기법을 이용하여 x-선 조사로 유도된 focus 형성에 관여하는 유전자 발현 패턴 변화를 연구해보고자 하였다. 주성분 분석과 admixture 분석 결과는 focus 형성 여부에 따라 전사체 패턴이 달라짐을 보여주었으며 focus 형성을 유도하는 인자를 차별발현 유전자 분석을 이용하여 규명하고자 하였다. X-선이 조사되지 않은 컨트롤군과 Non-focus군의 유전자 발현량과 패턴의 차이가 대조되었으며 그 결과, Non-focus군은 엑스레이로 유도된 형질변형 과정의 가장 초기단계에 있다는 것을 설명할 수 있었다. 또한, Non-focus군의 전사체 기질이 컨트롤군 보다는 Focus군에 더 가깝다는 것도 밝혀졌다. 특히, Non-focus군에서는 DNA 손상반응 유전자로 잘 알려진 *Cdkn1a*가 엑스레이로 인한 DNA 손상에 대한 반응 작용으로 발현량이 매우 높게 올라간 것을 관찰할 수 있었다. 그리고 focus 형성에 기여하는 인자로 생각되는 TGF-β 유전자들이 Non-focus군에서 발현이 떨어진 것을 관찰할 수 있었다. 게다가 DNA 회복 시스템 관련 유전자로 알려진 *Atm*, *Atr*, *Brca1*, *Brca2*, 그리고 *Chek1* 유전자들의 발현이 Focus군에서 매우 높게 올라간 것을 알 수

있었다. DNA 회복 시스템이 Focus군에서 전반적으로 많이 올라갔지만 특히, error-prone DNA 회복 시스템이 Focus군에서 높게 올라간 것을 볼 수 있었다. 세포가 외부 자극에 의해 손상을 입었을 경우, error-free 시스템이 먼저 작동을 하고 error-free 시스템이 과부하가 되었을 경우 error-prone 시스템이 가동이 된다. 본 연구 결과에서도 error-prone 시스템이 올라간 것을 관찰할 수 있었는데 error-prone 시스템에 의해 손상된 DNA가 회복이 되어 결국 제대로 회복이 되지 않고 마침내 focus를 형성하는 것으로 생각이 된다. Focus 형성에 따른 유전체 변화는 뚜렷하지 않았으나 전사체의 변화는 매우 극명하게 나타나는 것을 유전자 발현 패턴 변화 연구를 통해 알 수 있었다. 이는 암 유전체 진화적인 관점에서 볼 때, 유전체의 변화에 앞서 전사체의 변화가 1차적으로 먼저 일어나는 경우도 있다는 것을 시사한다. 마지막으로 선행연구 중 종양 발생과 줄기세포가 밀접한 연관이 있다는 보고가 있었다. 이를 바탕으로 Focus가 stem cell-like 성질이 있는지를 관찰하였고 그 결과, Focus는 stemness가 있는 것을 확인하였다. 또한 up-regulated된 *Myc* 유전자가 세포를 reprogramming 할 수 있다는 보고가 있었으며 *Myc*으로 유도된 종양은 유전체의 변화가 크게 없다고 하였다. Focus는 *Myc*의 발현이 올라가 있었으며 유전체의 변화 역시 크지 않았음으로 그 결과가 선행연구와 일치한다. 본 연구의 결과를 종합해보면 Little이 제시한 two-step 형질전환 이론에서 두 번째 단계에 관여하는 인자는

두 가지가 있는데 첫 번째로는 *Tgfβ* 유전자의 발현이 떨어지는 것이고 두 번째로는 올라간 error-prone DNA 회복 시스템이다. 아주 이른 시기의 종양 세포에서처럼 Non-focus군에서 *Tgfβ* 유전자의 발현이 떨어짐으로써 Non-focus 세포가 Focus 세포로 형질전환이 일어나는 것에 기여를 하였고 error-prone DNA 회복 시스템이 가동됨에 의해 제대로 복구 되지 않은 세포들이 누적이 되어 결국 focus를 형성하였다. 그리고 마지막으로 Focus가 갖고 있는 stem cell-like 성질과 *Myc*으로 유도된 oncogenic reprogramming으로 인해 transformed cell인 Focus가 malignancy를 갖게 된다고 생각할 수 있다. 따라서 Little의 two-step 형질전환 이론을 정확히 밝혀냄으로써 본 연구를 통해 종양 발생에 대한 이해도가 높아질 것으로 예상된다.

주요어: 종양발생, 형질변형, X-선 조사, 차세대 염기서열 분석, 전사체 분석, 전장 유전체 분석, error-prone DNA 회복 시스템, 암 줄기세포, oncogenic reprogramming

학번: 2015-31237